The EMBO Journal Vol.16 No.17 pp.5235–5246, 1997

RNA recognition by the joint action of two nucleolin RNA-binding domains: genetic analysis and structural modeling

Philippe Bouvet[1], Chaitanya Jain[2], Joel G.Belasco[2], François Amalric and Monique Erard

Laboratoire de Biologie Moléculaire Eucaryote, Institut de Biologie Cellulaire et de Génétique du CNRS, UPR 9006, 118 route de Narbonne, 31062 Toulouse Cedex, France and [2]Skirball Institute of Biomolecular Medicine, New York University Medical Center, 540 First Avenue, New York, NY 10016, USA

[1]Corresponding author
e-mail: Bouvet@ibcg.biotoul.fr

The interaction of nucleolin with a short stem–loop structure (NRE) requires two contiguous RNA-binding domains (RBD 1+2). The structural basis for RNA recognition by these RBDs was studied using a genetic system in *Escherichia coli*. Within each of the two domains, we identified several mutations that severely impair interaction with the RNA target. Mutations that alter RNA-binding specificity were also isolated, suggesting the identity of specific contacts between RBD 1+2 amino acids and nucleotides within the NRE stem–loop. Our data indicate that both RBDs participate in a joint interaction with the NRE and that each domain uses a different surface to contact the RNA. The constraints provided by these genetic data and previous mutational studies have enabled us to propose a three-dimensional model of nucleolin RBD 1+2 bound to the NRE stem–loop.

Keywords: *Escherichia coli* genetics/nucleolin/RNA-binding domain/RNA-binding specificity/structural modeling

Introduction

The consensus RNA-binding domain (CS-RBD), also called RNA recognition motif (RRM), is found in a large number of RNA-binding proteins involved in all aspects of post-transcriptional regulation (for recent reviews, see Burd and Dreyfuss, 1994a; Nagai *et al.*, 1995). These proteins often contain one to four CS-RBDs (Kenan *et al.*, 1991; Birney *et al.*, 1993). The three-dimensional structure of this conserved 70–90 amino acid RBD has been determined for only a few CS-RBD proteins. The best characterized of these is the spliceosomal protein U1A, which binds to hairpin II of U1 snRNA (Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990; Howe *et al.*, 1994; Oubridge *et al.*, 1994) and to a structurally related RNA element within the 3′-untranslated region of its own pre-mRNA (Allain *et al.*, 1996; Gubser and Varani, 1996; Jovine *et al.*, 1996). X-ray crystallographic and NMR studies of the N-terminal CS-RBD of the U1A protein have revealed that this domain comprises a four-stranded antiparallel β-sheet flanked on one side by two α-helices (Nagai *et al.*, 1990; Hoffman *et al.*, 1991). This structure is shared by other CS-RBDs (Görlach *et al.*, 1992; Wittekind *et al.*, 1992; Garret *et al.*, 1994; Lee *et al.*, 1994).

Among the most conserved features of the CS-RBD is the presence of two sequence motifs of eight and six amino acids (RNP-1 and RNP-2, respectively). Located on two adjacent β-strands (β2 and β3), these conserved motifs include aromatic residues thought to contact the RNA target directly (Merril *et al.*, 1988; Jessen *et al.*, 1991). This prediction was confirmed by the crystal structure of the U1A protein bound to U1 hairpin II, which revealed that the RNP-1 and RNP-2 motifs interact extensively with nucleotides of the RNA loop and that the polypeptide turn that links the β2 and β3 strands (turn 3) protrudes through the RNA loop (Oubridge *et al.*, 1994).

With the development of *in vitro* selection techniques (SELEX) (Tuerk and Gold, 1990; Tsai *et al.*, 1991), the RNA-binding specificities of a growing number of CS-RBD-containing proteins have been determined: hnRNP A1 (Burd and Dreyfuss, 1994b; Shamoo *et al.*, 1994; Abdul-Manan *et al.*, 1996), hnRNP C (Görlach *et al.*, 1994a), Sxl (Inoue *et al.*, 1992; Sakashita and Sakamoto, 1994), ASF/SF2 (Caceres and Krainer, 1993; Tacke and Manley, 1995), poly(A)-binding protein (Görlach *et al.*, 1994b; Kühn and Pieler, 1996), HuD (Chung *et al.*, 1996), HuC (Abe *et al.*, 1996) and nucleolin (Ghisolfi *et al.*, 1996). In some cases (U1A, hnRNP C), a single CS-RBD plus adjacent sequences is responsible for the RNA-binding specificity of the protein (Görlach *et al.*, 1994a; Oubridge *et al.*, 1994; Avis *et al.*, 1996), whereas in other cases (hnRNP A1, Sxl, ASF/SF2, HuD, HuC, nucleolin) it appears that multiple CS-RBDs cooperate in recognizing a shared RNA target (Burd and Dreyfuss, 1994b; Shamoo *et al.*, 1994; Kanaar *et al.*, 1995; Tacke and Manley, 1995; Abe *et al.*, 1996; Chung *et al.*, 1996; Serin *et al.*, 1997). Cooperation between two CS-RBDs was proposed for this latter group of proteins because each individual CS-RBD alone could not reproduce the binding specificity and affinity of the full-length protein (Burd and Dreyfuss, 1994b; Tacke and Manley, 1995; Serin *et al.*, 1997). The requirement for two CS-RBDs was demonstrated further by the fact that mutating conserved aromatic residues within the RNP-1 motif of each CS-RBD of hnRNP A1 (Merrill *et al.*, 1988; Mayeda *et al.*, 1994), ASF/SF2 (Caceres and Krainer, 1993; Zu and Manley, 1993) and nucleolin (Serin *et al.*, 1997) drastically impaired interaction with the RNA target. It is unclear how the two RBDs of these proteins interact specifically with a shared RNA target; nor has the role played by each domain been determined for any of these proteins.

An interesting feature of these dual-RBD interactions is that the two individual domains involved in RNA recognition are separated by a limited number of amino acids: 10 residues for Sxl and HuD, 12 for nucleolin, 17

for hnRNP A1 and 32 for ASF/SF2. This short distance between the two CS-RBDs may have two major consequences for the interaction with the RNA target. First, it should restrict the positioning of one CS-RBD in relation to the other, and secondly the close proximity of the two CS-RBDs should favor the interaction of each domain with the same RNA molecule (Shamoo *et al.*, 1994, 1995).

Nucleolin contains four CS-RBDs (Lapeyre *et al.*, 1987; Srivastava *et al.*, 1989) and interacts specifically with an RNA hairpin called the NRE (Ghisolfi *et al.*, 1996). A detailed deletion and mutational analysis (Serin *et al.*, 1997) revealed that the first two CS-RBDs (RBD 1+2) are necessary and sufficient for the specific, high-affinity interaction with the RNA target. To gain insight into the mechanism of the interaction between this dual-RBD protein and its RNA target, we have used a genetic strategy based on the repression of *lacZ* translation by a heterologous RNA-binding protein expressed in *Escherichia coli* (Jain and Belasco, 1996). Using this genetic strategy, we have identified amino acid mutations within each of the two nucleolin CS-RBDs that completely abolish interaction with the NRE. Furthermore, we have isolated protein suppressor mutations that partially compensate for the deleterious effect of mutations within the NRE. These different constraints make it possible to suggest a model for the interaction of the two nucleolin RBDs with their shared RNA target.

Results

Specific interaction of two nucleolin RBDs with their RNA target in E.coli

We previously identified a short RNA stem–loop (the NRE) as the high-affinity RNA target of nucleolin (Ghisolfi *et al.*, 1996). Specific interaction of nucleolin with the NRE is mediated by its first two CS-RBDs (RBD 1+2) (Serin *et al.*, 1997). The full integrity of these two RBDs is necessary and sufficient to account for the RNA-binding specificity of nucleolin. Each of these two domains appears to be involved in direct interaction with the RNA target, since mutating conserved aromatic residues within each RNP-1 motif drastically impairs NRE binding. To identify amino acids involved in this interaction, we made use of a genetic system recently developed by Jain and Belasco (1996). This system is based on the translational repression of *lacZ* by a heterologous RNA-binding protein that sterically hinders ribosome binding by binding to an RNA target sequence inserted a few nucleotides upstream of the *lacZ* Shine–Dalgarno element.

To use this system, the minimal 18 nucleotide RNA sequence required for nucleolin binding (Ghisolfi *et al.*, 1996; Serin *et al.*, 1996) was first introduced 11 nucleotides upstream of the Shine–Dalgarno element of a *lacZ* reporter construct (Figure 1B). A *lacZ E.coli* strain (WM1/F′) transformed with the resulting plasmid (pLacNRE) synthesizes high levels of β-galactosidase, producing blue colonies on X-Gal indicator plates. Expression of *lacZ* from this plasmid was strongly repressed upon transformation with a second plasmid (pRBD1+2; Figure 1A) that encodes a truncated protein comprising the two nucleolin RBDs sufficient for NRE binding *in vitro*. β-Galactosidase synthesis in these double transformants was reduced by a factor of 36 (the repression ratio) compared with isogenic cells transformed with the same reporter plasmid and plasmid pACYC184, which does not encode nucleolin (Table I).

To confirm that the translational repression of the NRE-*lacZ* transcript was due to a specific interaction of the RBD 1+2 protein with the NRE, we transformed *E.coli* cells containing the pRBD1+2 plasmid with a series of NRE-*lacZ* reporter plasmids bearing mutations in the NRE (pLacM6, M10, M11, M12, M13) (Figure 1C). These RNA point mutations previously had been shown to severely impair nucleolin binding *in vitro* (Ghisolfi *et al.*, 1996; Serin *et al.*, 1997), and they resulted in a substantial increase in β-galactosidase activity in *E.coli* due to poor binding of the RBD 1+2 protein to the mutant reporter transcripts (Table I). An additional control was performed using two mutant forms of the RBD 1+2 protein, R1LL and R2LL. These mutants contain pairs of conservative amino acid substitutions in the RNP-1 motif of RBD 1 and RBD 2 (R1LL: F43L and Y45L; R2LL: I125L and Y127L) that previously have been shown to abolish NRE binding *in vitro* (K_d >10 μM, Serin *et al.*, 1997). No repression of *lacZ* translation was observed when these two protein mutants were expressed in cells containing the wild-type NRE-*lacZ* reporter plasmid (Table II), indicating that these mutant proteins fail to bind the NRE in *E.coli*. A similar lack of repression was observed for R1LL and R2LL in cells producing various mutant NRE-*lacZ* reporters (M6, M10, M11, M12 and M13).

To determine how well translational repression of the NRE-*lacZ* transcript correlates quantitatively with binding affinity, we used gel-shift analysis to measure the dissociation constant (K_d) of several RBD 1+2 proteins and various NRE RNAs (Figure 2A and data not shown). A T7 RNA polymerase promoter located a short distance upstream of the NRE sequence in pLacNRE was used for *in vitro* synthesis of labeled RNAs. It was particularly important in these studies to measure the affinity of the RBD 1+2 protein for the wild-type NRE within the context of the *lacZ* reporter sequence. The wild-type RBD 1+2 protein was found to bind the NRE-*lacZ in vitro* transcript with a high affinity (K_d of 20 ± 5 nM) (Figure 2A) identical to that previously measured for the NRE in a different RNA context (Ghisolfi *et al.*, 1996; Serin *et al.*, 1997). Similarly, the affinity of this protein for various NRE mutants (Figure 2A and data not shown) and of the R1LL protein mutant for the wild-type NRE were also unchanged in this new context. These results show that the affinity of the RBD 1+2 protein for the NRE is context independent. Moreover, for K_d values between 10 nM and 10 μM, there is a remarkably good correlation between the binding affinity measured *in vitro* and the degree of *lacZ* translational repression observed in *E.coli* (Figure 2B).

Together, these results demonstrate that the specific interaction of the RBD 1+2 protein and the NRE can be faithfully reproduced in *E.coli*. This knowledge enabled us to use this genetic system to investigate the interaction of this dual-RBD protein with its RNA target.

Identification of nucleolin amino acids involved in RNA binding

We first used this genetic system to identify amino acids whose mutation abolishes the ability of the RBD 1+2

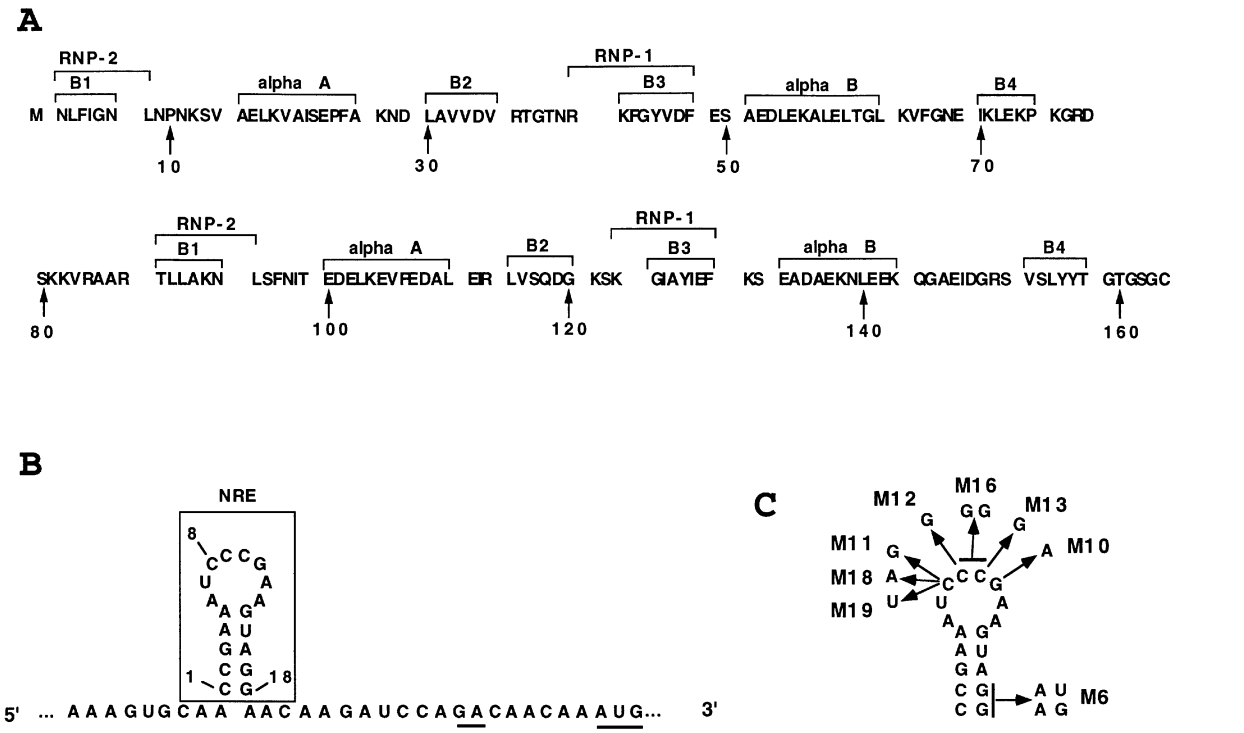

Fig. 1. The RBD 1+2 protein and its RNA target. (**A**) The sequence of the RBD 1+2 protein produced by the pRBD1+2 plasmid is indicated. Amino acid residues are numbered from the methionine introduced before the first residue of the RNP-2 motif of CS-RBD1. Elements of secondary structure and the conserved RNP-1 and RNP-2 sequence motifs are indicated. (**B**) Sequence of the nucleolin recognition element (NRE) in the context of the NRE-*lacZ* transcript of pLac-NRE. The boxed 18 nucleotide stem–loop represents the minimal nucleolin-binding site determined by SELEX (Ghisolfi *et al.*, 1996). Underlined GA and AUG nucleotides indicate the rudimentary Shine–Dalgarno element and the initiation codon of *lacZ*. (**C**) Representation of the different NRE mutants used in this study. Each mutation involved one or more substitutions at the indicated sites. Mutation M6 disrupts the secondary structure of the NRE (Ghisolfi *et al.*, 1996).

protein to bind the NRE. Translational repression of the NRE-*lacZ* transcript by the RBD 1+2 protein inhibits β-galactosidase synthesis, giving rise to white colonies when cells are grown on X-Gal plates. In contrast, cells that produce the R1LL and R2LL proteins, which are unable to bind the wild-type NRE, generate blue colonies. This phenotypic difference provides a convenient basis to screen for protein mutants deficient in RNA binding.

Random PCR mutagenesis was performed on RBD 1+2 cDNA, which was then subcloned back into the parent plasmid to generate the mutagenized plasmid library pRBD1+2M. The *E.coli* WM1/F′ cells were co-transformed with this library and the pLacNRE reporter plasmid and plated on X-Gal plates. About half of the resulting colonies were blue, indicating that these cells produced a mutant form of the RBD 1+2 protein unable to interact with the NRE. However, Western blot analysis revealed that only ~30% of the blue colonies expressed the RBD 1+2 protein variant at a level comparable with wild-type (data not shown); the remaining cells expressed either a truncated protein or no protein at all. Many of these mutations might impair proper protein folding, resulting in an accelerated rate of degradation (Pakula *et al.*, 1986).

Plasmid DNA was purified from cells expressing RBD 1+2 variants that were present at a wild-type concentration yet deficient for *lacZ* translational repression. DNA sequencing was then performed to identify the mutation responsible for this loss of function (Table III). The location of these mutations was interesting in two respects. First, amino acid mutations that abolish RNA binding could be found in either RBD 1 or RBD 2, demonstrating that both RBDs participate in RNA binding. Secondly, the mutations did not appear to be randomly distributed: in RBD 1, the RNP-2 and RNP-1 motifs and the β2–β3 loop were highly affected, whereas in RBD 2, helix A and the RNP-1 motif were the major sites for the mutations. The localization of these mutations was not a consequence of the PCR-based mutagenesis method, since sequencing of random clones did not show any preference for mutation of these domains (data not shown).

Table I. Specific interaction of the RBD 1+2 protein with NRE RNA in *E.coli*

pLac plasmid	β-Galactosidase activity		Repression ratios	K_d (nM)
	pACYC184	pRBD 1-2		
pLacNRE	291 ± 9	7.6 ± 0.6	36.3 ± 2.2	20 ± 5
pLacM6	229 ± 10	105.5 ± 5	2.2 ± 0.1	1000 ± 200
pLacM10	345 ± 3	179.5 ± 4.5	1.9 ± 0.1	3000 ± 500
pLacM11	148 ± 4	68.7 ± 0.65	2.1 ± 0.02	3000 ± 500
pLacM12	232 ± 11	28.4 ± 1.3	8.0 ± 0.1	250 ± 50
pLacM13	122 ± 6	101.5 ± 1.5	1.2 ± 0.1	6000 ± 1000

Escherichia coli strain WM1/F′ was co-transformed with pRBD1+2 or parent plasmid pACYC184 and with a pLacNRE reporter plasmid containing the wild-type nucleolin recognition element or NRE mutant M6, M10, M11, M12 or M13. β-galactosidase activity in the resulting strains was quantified as described in Materials and methods. Repression ratios were determined by dividing the β-galactosidase activity in cells containing pACYC184 by the β-galactosidase activity in cells containing the pRBD1+2 plasmid.

Table II. Interaction of mutated RBD 1+2 proteins with wild-type and mutated NRE stem–loops

pLac plasmid	Repression ratios		
	pRBD 1-2	pR1LL	pR2LL
pLacNRE	38.5 ± 2.0	1.12 ± 0.10	0.74 ± 0.13
pLacM6	2.0 ± 0.2	1.01 ± 0.06	0.74 ± 0.13
pLacM10	2.0 ± 0.2	1.01 ± 0.03	0.87 ± 0.05
pLacM11	2.1 ± 0.7	1.01 ± 0.02	0.89 ± 0.08
pLacM12	7.6 ± 0.4	0.98 ± 0.01	0.97 ± 0.03
pLacM13	1.4 ± 0.3	1.08 ± 0.07	0.90 ± 0.14

Previous studies (Serin *et al.*, 1997) have shown that mutation of conserved aromatic residues within the RNP-1 motif of RBD 1 (R1LL) and RBD 2 (R2LL) abolishes NRE binding *in vitro*. cDNA encoding the mutated protein was substituted for the corresponding fragment in pRBD1+2 to give pR1LL and pR2LL (see Materials and methods for details). The resulting plasmids were co-transformed with various pLac reporter plasmids. Repression ratios were determined from the β-galactosidase activity with or without the RBD 1+2 protein, as in Table I.

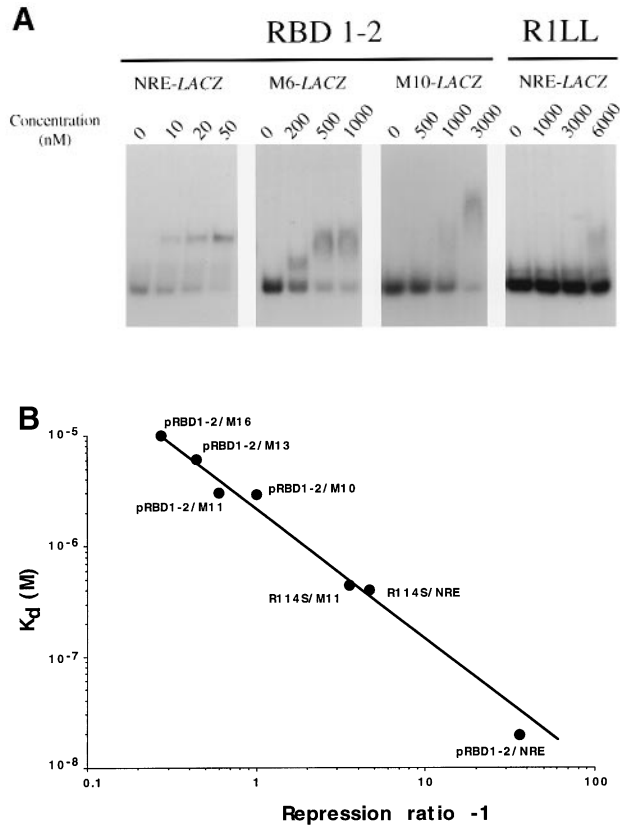

Fig. 2. (**A**) Representative gel-shift analysis of the interaction of a wild-type or mutant RBD 1+2 protein with a wild-type or mutant NRE. ^{32}P-Labeled RNA was synthesized by *in vitro* transcription of the corresponding pLacNRE plasmid and incubated with different concentrations of the protein. Free RNA was separated from the RNA–protein complex in an 8% polyacrylamide gel under non-denaturing conditions. (**B**) Correlation between the repression ratios measured in *E.coli* and dissociation constants (K_d) measured *in vitro*. The results obtained in Table I and with other protein and RNA mutants (see below, e.g. in Table IV) were plotted on this graph (correlation coefficient of 0.968). Note that 1 must be subtracted from the repression ratios prior to their quantitative comparison, as a repression ratio of 1 indicates the absence of detectable binding.

Table III. Identification of amino acid residues important for the binding of RBD 1+2 to the NRE

Location	Mutations
RBD 1	
RNP-2 motif	F4L, N7T, L8R, L8P
loop 3	R36G, N40K
RNP-1 motif	F48G
RBD 2	
α A helix	F107Y, F107L, D109N, L111W, L111S (2)
RNP-1 motif	I128S, F130S (3), F130L

RBD 1+2 protein variants unable to bind the wild-type NRE were isolated, and the mutated amino acids were identified by DNA sequencing. Mutants that were independently isolated more than once are indicated. In cells containing the reporter plasmid pLacNRE, the activity of β-galactosidase in the presence of any of these RBD 1+2 variants is identical to the activity of β-galactosidase in the presence of plasmid pACYC184 (data not shown), indicating that these RBD 1+2 variants fail to bind the NRE.

It is not surprising that mutating the RNP-1 and RNP-2 motifs of RBD 1+2 impairs NRE binding, as these motifs have been implicated in RNA binding by other CS-RBD proteins. Likewise, in the complex of the U1A protein and its RNA hairpin target, the β2–β3 loop protrudes through the RNA loop and plays an important role in the specificity of this interaction (Scherly *et al.*, 1990; Bentley and Keene, 1991). Thus, the two mutations found in the corresponding RBD 1 loop (R36G, N40K) may suggest a similar role for this loop in the interaction of RBD 1+2 with the NRE stem–loop (see Discussion). The mutations in helix A of RBD 2 were unexpected, as the corresponding helix of U1A is not in close proximity to the RNA in that RBD–RNA complex (Oubridge *et al.*, 1994).

Identification of altered-specificity RBD 1+2 variants

These loss-of-function mutations provide interesting information as to the amino acids that are important for RNA binding. However, they do not reveal whether the mutated amino acid directly contacts the nucleic acid, since a lack of RNA binding could also result from mutations that modify the structure of the RBD.

To understand how the RBD 1+2 protein interacts with the NRE, specific contacts between amino acids and nucleotides must be identified. To access this kind of information, we used our genetic system to screen for mutations in RBD 1+2 that alter the specificity of this protein and enable it to bind with increased affinity to mutant NREs. Using this gain-of-function strategy, we hoped to acquire more specific information about protein–RNA interactions than could be obtained using the loss-of-function strategy described above.

To this end, the pRBD1+2M plasmid library was introduced into a set of *E.coli* strains that each contained a mutant reporter plasmid (pLacM6, M10, M11, M13 or M16). Colonies that appeared less blue in color than control colonies containing the wild-type pRBD1+2 plasmid were tested again to confirm that they contained a reduced level of β-galactosidase activity. Each of the mutant pRBD1+2 plasmids that survived this second screen was purified and sequenced to identify the mutated amino acid (Table IV). Interestingly, the affected amino acids were located

Table IV. Repression ratios for RBD 1+2 variants with NRE mutants

	wt NRE UCCCGAA	M6 UCCCGAA	M10 UCCCAAA	M11 UGCCGAA	M12 UCGCGAA	M13 UCCGGAA	M16 UCGGGAA
wt pRBD 1-2	36.3 ± 2.2	2.0 ± 0.2	2.0 ± 0.3	2.1 ± 0.7	7.7 ± 0.4	1.4 ± 0.3	1.3 ± 0.1
N2T	38.0 ± 4.0	**6.3 ± 0.7**	**5.0 ± 0.2**	**9.5 ± 0.8**	5.9 ± 0.8	**2.0 ± 0.1**	1.5 ± 0.1
R114S	5.8 ± 0.5	1.0 ± 0.2	0.9 ± 0.02	**4.5 ± 0.6**	1.7 ± 0.2	0.8 ± 0.1	1.1 ± 0.1
Y45F	35.0 ± 1.3	**12.8 ± 0.8**	**5.0 ± 0.8**	**11.4 ± 1.5**	8.7 ± 0.2	**3.5 ± 0.4**	**4.0 ± 0.6**
Y45F/Y127F	37.9 ± 0.4	**4.2 ± 0.5**	**3.4 ± 0.6**	**6.2 ± 1.5**	**18.2 ± 0.3**	**2.2 ± 0.5**	**4.6 ± 1.5**
E108D	18.4 ± 1.2	**7.4 ± 1.0**	**4.7 ± 0.2**	**7.4 ± 2.5**	**15.8 ± 1.4**	**2.6 ± 0.01**	**1.8 ± 0.1**

A plasmid library (pRBD1+2M), randomly mutagenized in the RBD 1+2 gene, was co-transformed with one of six pLacNRE mutants: M6, M10, M11, M12, M13 or M16. Colonies that were lighter blue in color than control colonies containing the wild-type plasmid (pRBD1+2) were subjected to further analysis (see Materials and methods), including sequencing of the RBD 1+2 gene. RBD 1+2 mutants N2T and E108D were identified as suppressors of NRE mutant M6; R114S as a suppressor of M11; Y45F as a suppressor of M6, M10 and M16; and Y45F/Y127F as a suppressor of M16. Isolated plasmids encoding the different RBD 1+2 variants were retransformed with the other pLacNRE mutants, and the β-galactosidase activity determined for each plasmid combination. Repression ratios (R) in boldface indicate binding affinities at least twice as high as observed for the wild-type RBD 1+2 protein with the same RNA [$(R_i-1)/(R_j-1) > 2$].

in both RBDs (N2T and Y45F in RBD 1; E108D, R114S and Y127F in RBD 2).

The binding specificity of these RBD 1+2 suppressor mutants was then examined by testing their ability to repress the translation of various NRE-*lacZ* RNA mutants. These quantitative measurements of *lacZ* expression allowed the RBD 1+2 variants to be classified into three groups. In the first group, comprising the N2T, Y45F and Y45F/Y127F variants, the protein mutations each enhance translational repression of a number of NRE-*lacZ* variants, yet the level of repression of wild-type NRE-*lacZ* translation is the same as that observed with the wild-type RBD 1+2 protein (Table IV). To show that the increased repression of NRE-*lacZ* mutants by the Y45F variant is really a consequence of an increase in binding affinity, this mutant protein was purified from *E.coli* and studied in gel-shift experiments. These binding studies (Figure 3) confirmed that the Y45F mutant binds the wild-type NRE with the same affinity as the wild-type RBD 1+2 protein (K_d of 20 nM) and has a significantly higher affinity than the wild-type protein for mutated NREs (see, for example, M6-*lacZ* in Figure 3). That all three of these RBD 1+2 suppressor mutants contain amino acid substitutions within the RNP-1 and RNP-2 motifs of RBD 1 strongly suggests that these protein segments are involved in determining the binding specificity of the RBD 1+2 protein. The Y45F variant in particular was independently isolated many times in suppressor screens involving three different NRE-*lacZ* mutants (M10, M6, M16), indicating the importance of this amino acid residue for the interaction of the RBD 1+2 protein with the NRE.

A second class of RBD 1+2 suppressor mutants was more effective at repressing the translation of multiple NRE-*lacZ* variants but less effective at repressing the wild-type NRE-*lacZ* reporter transcript. Thus, the E108D mutation significantly increases the repression of several different reporter mutants but represses wild-type NRE-*lacZ* expression only half as well as the wild-type RBD 1+2 protein.

The third type of RBD 1+2 suppressor mutation that was isolated (R114S) specifically increased translational repression of a single NRE-*lacZ* mutant and significantly impaired repression of the wild-type reporter. This substitution of a serine residue for Arg114 caused a 2-fold enhancement in repression of the M11-*lacZ* transcript,

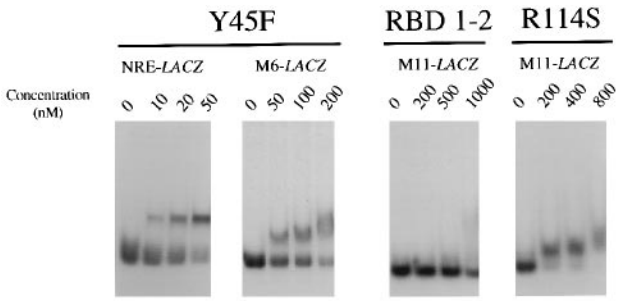

Fig. 3. Representative gel-shift analysis of the binding of the wild-type or mutated RBD 1+2 protein to wild-type or mutated NRE RNA. Mutated proteins were produced in *E.coli*, purified to homogeneity and studied in gel-shift experiments as described in Figure 2A.

while reducing translational repression of the wild-type NRE-*lacZ* transcript by a factor of six. In contrast, R114S repression ratios close to 1 (no repression) were measured for the M6, M10, M13 and M16-*lacZ* mutants, indicating that this protein variant had completely lost the ability to interact with these other NRE mutants.

To confirm that the enhanced repression of the M11 reporter was the result of an increase in binding affinity, the R114S protein was purified from *E.coli* and studied *in vitro* by gel-shift analysis (for a representative gel, see Figure 3). These measurements showed that the R114S variant binds M11 RNA about five times more tightly than does the wild-type protein (K_d of 400 nM, versus 2000 nM for wild-type RBD 1+2; Figure 3). Moreover, this amino acid substitution reduces the affinity of the RBD 1+2 protein for the wild-type NRE by about a factor of 15 (K_d of 300 nM, versus 20 nM for the wild-type protein) (data not shown). These altered binding characteristics suggest that, in the wild-type protein, Arg114 may be involved in recognition of the cytosine residue (C8) that is replaced by guanosine in the M11 mutant.

Mutational analysis of amino acid residue R114

Because a single point mutation within codon 114 can give rise to only five different amino acid substitutions, it was possible that Arg114 mutations other than R114S might better compensate the M11 RNA mutation. To examine more comprehensively the importance of amino

acid residue 114 in NRE recognition, we completely randomized the RBD 1+2 codon corresponding to this residue, thereby creating the plasmid library pR114Lib. This library was used to transform cells carrying one of four different NRE-*lacZ* reporter plasmids: wild-type, M11, M18 or M19, which have a cytosine, guanosine, adenosine or uridine, respectively, at position 8 of the NRE (see Figure 1C). In each case, RBD 1+2 variants able to repress *lacZ* expression (white colonies) were identified. Plasmids from these colonies were isolated and sequenced to reveal the identity of amino acid 114.

A significant bias for certain amino acids was observed with each of the four RNA targets (Table V), suggesting again that amino acid 114 is important for the recognition of nucleotide 8 of the NRE. With the wild-type NRE (C8), all 17 independent RBD 1+2 clones that were sequenced had an arginine residue at position 114 (all six possible arginine codons were represented; data not shown). A strong bias in favor of a basic amino acid was also observed for the M19 mutant (U8) (arginine or lysine in all of the repressing RBD 1+2 clones). In contrast, threonine, serine, asparagine and aspartate residues were favored for M11 (G8), and threonine, serine, arginine, aspartate and glutamine residues were favored for M18 (A8) (Table V).

Table V. Screening for amino acid substitutions at position 114 that can suppress RNA mutations at C8

LacZ plasmid	Selected amino acid	Frequency (%)
wt NRE UCCCGAA	R	100
M11 UGCCGAA	T	53
	S	30
	N	12
	D	5
M18 UACCGAA	S	30
	T	30
	R	13
	D	13
	Q	13
M19 UUCCGAA	R	85
	K	15

Saturation mutagenesis was performed on amino acid 114 of RBD 1+2. Cells were co-transformed with the resulting plasmid library (pR114Lib) and pLacNRE, pLacM11, pLacM18 or pLacM19. Transformants containing an RBD 1+2 variant that could bind the co-resident reporter mRNA were identified by their white or light blue colony phenotype on X-Gal plates. The pRBD1+2 plasmid in these cells was isolated and sequenced. Indicated for each NRE is the frequency with which various amino acids appeared at position 114 among the isolates. The number of individual RBD 1+2 clones that were sequenced was 15 for the wild-type NRE, 17 for M11, seven for M18 and seven for M19.

To determine the specificity of RBD 1+2 variants with threonine, asparagine, aspartate, glutamine or lysine substitutions at position 114, plasmids encoding these protein variants were introduced into *E.coli* strains containing reporter plasmids with NRE mutations at various positions. In each case, the degree of translational repression was determined by spectrophotometric measurements of β-galactosidase activity (Table VI). All of these RBD 1+2 mutations increase translational repression of M11-*lacZ* (G8) and reduce repression of wild-type NRE-*lacZ* (C8). None of the R114 variants is more effective than the wild-type RBD 1+2 protein at repressing translation of reporter mRNAs bearing other NRE mutations, and in most cases the degree of repression is significantly less. The RBD 1+2 variant most specific for M11 is R114D. On the basis of the observed correlation between relative binding affinity and repression ratio (Figure 2B), we estimate that this amino acid substitution causes a 4-fold increase in the affinity of RBD 1+2 for M11 RNA (calculated K_d = 500 nM) while reducing the affinity of the protein for the wild-type NRE by a factor of ~100 (calculated K_d = 2000 nM). The greater affinity of this mutant protein for the M11 stem–loop versus the wild-type NRE hairpin makes R114D a true altered-specificity variant. Together, these findings support the conclusion that Arg114 of RBD 1+2 plays a key role in the recognition of NRE nucleotide C8, suggesting a direct interaction between these two residues.

Discussion

A genetic system to study the interaction of a dual-RBD protein with its RNA target

Nucleolin, a major non-ribosomal nucleolar protein, interacts specifically with an RNA stem–loop structure (NRE) through its first two RNA-binding domains (RBD 1+2) (Ghisolfi *et al.*, 1996; Serin *et al.*, 1996, 1997). We have successfully used a rapid genetic screening procedure in *E.coli* (Jain and Belasco, 1996) to study the interaction of this dual-RBD protein and its RNA target. This genetic approach is based on the ability of RBD 1+2 to specifically repress translation of a *lacZ* reporter transcript by binding

Table VI. Repression ratios of Arg114 variants with NRE mutants

	wt NRE UCCCGAA	M6 UCCCGAA	M10 UCCCAAA	M11 UGCCGAA	M12 UCGCGAA	M13 UCCGGAA	M16 UCGGGAA	M18 UACCGAA	M19 UUCCGAA
wt pRBD 1-2 R114	36.3 ± 2.2	2.0 ± 0.2	2.0 ± 0.3	2.1 ± 0.7	7.6 ± 0.4	1.4 ± 0.3	1.3 ± 0.1	6.9 ± 0.7	12.5 ± 0.9
R114S	5.8 ± 0.5	0.9 ± 0.2	0.90 ± 0.02	**4.5 ± 0.6**	1.7 ± 0.2	0.9 ± 0.1	1.1 ± 0.1	10.7 ± 0.4	4.5 ± 0.8
R114T	4.1 ± 0.1	1.0 ± 0.2	1.1 ± 0.2	**4.7 ± 0.4**	3.8 ± 0.6	0.9 ± 0.1	0.89 ± 0.02	9.6 ± 0.8	3.6 ± 0.2
R114N	5.2 ± 0.5	1.1 ± 0.3	0.9 ± 0.2	**4.6 ± 0.6**	3.8 ± 0.3	0.9 ± 0.2	0.9 ± 0.1	10.2 ± 0.4	6.1 ± 0.3
R114D	1.6 ± 0.1	1.0 ± 0.1	0.8 ± 0.1	**4.5 ± 0.5**	1.2 ± 0.1	1.3 ± 0.1	1.01 ± 0.01	2.7 ± 0.6	1.5 ± 0.2
R114Q	2.5 ± 0.3	0.98 ± 0.01	1.1 ± 0.1	**3.15 ± 0.02**	1.3 ± 0.1	1.4 ± 0.1	1.11 ± 0.02	5.0 ± 0.3	2.2 ± 0.1
R114K	7.9 ± 0.7	1.0 ± 0.1	1.1 ± 0.2	**4.5 ± 0.4**	7.9 ± 0.3	1.5 ± 0.3	1.17 ± 0.03	6.8 ± 0.9	5.6 ± 0.1

Cells containing one of nine pLacNRE reporter plasmids were transformed with wild-type pRBD1+2, any of seven different pRBD 1+2 variants mutated at codon 114 or pACYC184 and β-galactosidase activity was determined. Repression ratios were calculated from β-galactosidase levels in each of the resulting strains. Repression ratios in boldface indicate binding affinities at least twice as high as observed for the wild-type RBD 1+2 protein with the same RNA [$(R_i-1)/(R_j-1) > 2$].

to an NRE hairpin inserted close to the ribosome-binding site. The degree of repression (the repression ratio) correlates quantitatively with the binding affinity of the protein for K_d values between 10 nM and 10 μM, validating the use of this genetic system to study the interaction between the RBD 1+2 protein and the NRE.

This genetic approach was first used to begin to identify amino acids important for binding. Our aim was not to identify every amino acid involved, but only to determine whether critical residues would be found in one or both RBDs. From the small number of defective RBD 1+2 mutants that were sequenced, it is clear that amino acids important for RNA binding are located in both RBDs (Table III). Among these amino acid residues are some that are potentially involved in direct contact with the RNA (Phe4, Asn2, Arg36, Asn40), as well as others more likely to be involved in maintaining the structural integrity of the RBDs (Leu8, Phe48, Phe107, Asp109, Leu111, Ile128, Phe130). Two of these important residues, Arg36 and Asn40, are located in an RBD 1 protein loop (the β2–β3 loop) that corresponds to one of the most variable regions among different RBD domains. The importance of this loop for RNA binding by nucleolin is consistent with previous studies of two other RBD proteins (U1A and U2B′) that implicate a corresponding protein loop in determining their binding specificity (Scherly *et al.*, 1990; Bentley and Keene, 1991).

To identify likely contacts between nucleolin protein residues and NRE nucleotides, we used our genetic system to screen for RBD 1+2 variants better able to bind mutated NRE stem–loops (Table IV and Figure 3). Although such gain-of-function mutations are expected to be rare in RNA-binding proteins, we were able in this manner to identify Arg114 as a key residue for the specificity of the interaction between RBD 1+2 and the NRE. Our genetic data clearly demonstrate that this arginine residue is required for tight binding to the wild-type RNA. Various amino acid substitutions at this position can alter the binding specificity of nucleolin with regard to the identity of the nucleotide at position 8 of the NRE, improving binding to the mutant RNA while impairing binding to the wild-type target. These nucleotide substitutions do not appear to cause any major rearrangements in RNA conformation, as judged by enzymatic probing (Ghisolfi *et al.*, 1996; P.Bouvet, unpublished data), suggesting that the altered specificity of the corresponding protein suppressor mutants is a consequence of a localized structural accommodation. These findings suggest that Arg114 lies in close proximity to C8 in the nucleolin–NRE complex. This amino acid residue is located in the protein loop connecting helix A and β strand 2 of RBD 2. As this protein loop is situated quite far from the intermolecular interface in the RNA complex of the N-terminal CS-RBD of U1A, our genetic data suggest that different CS-RBDs can use different protein surfaces to dock with their RNA targets.

Other interesting protein mutations improve binding to mutant NREs yet do not significantly affect binding to the wild-type RNA target. In these cases, the affected protein residues are located within the conserved RNP-2 and RNP-1 motifs of RBD 1 (Asn2, Tyr45) and RBD 2 (Tyr127). The Y45F variant, in particular, was isolated in screens for suppressors of three different NRE mutants, indicating that Tyr45 plays an important role in determining the RNA-binding specificity of nucleolin. This tyrosine residue potentially could interact with NRE nucleotides through a ring-stacking interaction, as observed for the corresponding RNP-1 residue of U1A (phenylalanine), and/or by hydrogen bonding (Oubridge *et al.*, 1994; LeCuyer *et al.*, 1996). Its replacement with phenylalanine would result in the loss of a single side-chain hydroxyl group. Further studies will be required to determine the structural basis for the increased affinity of the Y45F variant for many different NRE mutants.

Model for the interaction of a dual-RBD protein with an RNA stem–loop

It is becoming increasingly evident that the RNA-binding specificity of a large number of proteins that contain mutiple CS-RBDs results from cooperation between two RBDs (Burd and Dreyfuss, 1994b; Shamoo *et al.*, 1994; Kanaar *et al.*, 1995; Tacke and Manley, 1995; Chung *et al.*, 1996). So far, no high-resolution structure for the RNA complex of such a protein is available. The structural constraints suggested by our present genetic data and by additional binding studies with deletion and point mutants (Serin *et al.*, 1997) make it possible to propose a three-dimensional model for the nucleolin–NRE complex. In building the model, we have taken advantage of the homology of each nucleolin CS-RBD to the N-terminal CS-RBD of the splicesomal protein U1A, whose structure as a complex with hairpin II of U1 snRNA has been determined crystallographically (Oubridge *et al.*, 1994). A detailed description of the construction of this model can be found in Materials and methods.

The resulting model of the NRE complex of nucleolin RBD 1+2 (Figures 4 and 5) has a number of attractive features. In it, the interaction of RBD 1 with the NRE stem–loop bears a strong resemblance to the interaction of U1A with U1 hairpin II. Thus, RBD 1 residues Asn2, located at the beginning of the β strand 1 (RNP-2), and Arg36 and Asn40, located in the protein segment connecting β strands 2 and 3, are proposed to interact with the NRE loop (Figure 5A), consistent with their critical role in NRE binding and with the proximity of the corresponding U1A residues to the loop of U1 hairpin II. RBD 1 residues Phe4 (β1, RNP-2) and Tyr45 (β3, RNP-1) are proposed to stack with NRE nucleotides C10 and G11, respectively (Figure 4). These two aromatic residues are critical for RNA binding by nucleolin, as are the bases with which they are proposed to interact (Tables I–III, Figures 2 and 3; Ghisolfi *et al.*, 1996), and their homologs in U1A (Tyr13 and Phe56) are known to stack on adjacent loop nucleotides of U1 hairpin II (C10 and A11). In light of growing evidence that such base stacking interactions may be an evolutionarily conserved mechanism of CS-RBD–RNA interaction (Birney *et al.*, 1993; Nagai *et al.*, 1995), it seems reasonable that this recognition mechanism would apply to the interaction of RBD 1+2 with the NRE stem–loop.

In contrast, the proposed docking mode of RBD 2 with the NRE stem–loop is quite different. This would explain the distinct distribution of the critical RBD 2 residues identified in our genetic screens, none of which mapped to β strand 1 (RNP-2) or the β2–β3 loop. Instead, we identified altered-specificity mutations affecting residues

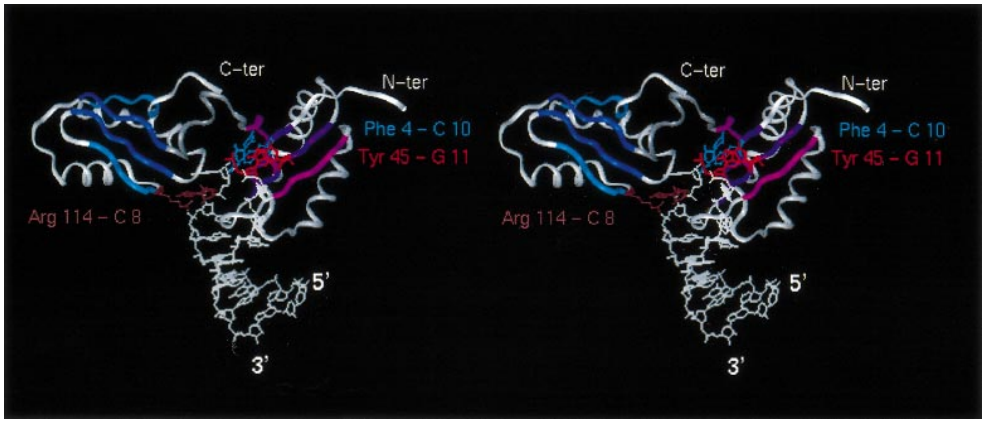

Fig. 4. Stereo-view of the computer model of the interaction between the RBD 1+2 protein and the NRE RNA stem–loop. RBD 1 β1, β3 strands are displayed in indigo and β2, β4 in magenta. RBD 2 β1, β3 strands appear in deep blue and β2, β4 strands in blue. Aromatic residues from RBD 1 are displayed in ball-and-stick mode, in blue for Phe4 and in red for Tyr45. The corresponding nucleotides from the NRE with which they are stacked appear in stick mode, in blue for C10 and red for G11. The specific contact between the Arg114 side chain and C8 has been color-coded in brown. Hydrogen atoms have been omitted for clarity.

in helix A (Glu108) and in the loop connecting helix A and β strand 2 (Arg114). A key feature of the model is that Arg114 of RBD 2 is shown contacting nucleotide C8 of the NRE loop, which would account for our genetic evidence that the identity of amino acid 114 determines the binding specificity of nucleolin at this RNA position. The relaxed binding specificity of the E108D mutant (Table IV) and the severe impediment to binding caused by several other mutations in RBD 2 helix A (Table III) may result from a repositioning of the adjacent helix A–β2 loop, which contains Arg114 and appears to be critical for RNA binding (Figure 5B). Our genetic evidence that RBD 2 binds RNA in a novel manner expands the repertoire of possible RNA-binding surfaces in CS-RBDs and raises the possibility that other dual-RBD proteins recognize RNA in a similar asymmetrical manner involving one RBD that binds in a U1A-like fashion and another RBD that binds in a distinct mode.

It is worth noting that in the proposed model for the RNA complex of nucleolin, the RNA-binding platforms of RBD 1 and RBD 2 (β sheets) are positioned in an antiparallel orientation on the same face of the protein. A similar relative orientation has been reported recently for two RBDs of hnRNP A1 in the absence of RNA (Shamoo *et al.*, 1997; Xu *et al.*, 1997).

Another interesting feature of the model is that, unlike their RBD 1 counterparts, the aromatic residues in the β1 (RNP-2) and β3 (RNP-1) strands of RBD 2 are prevented from stacking with NRE loop nucleotides by the proposed contact between Arg114 and C8, which would force these RBD 2 residues to lie some distance from the RNA–protein interface. Whether this RBD can employ its underutilized RNA-binding potential to interact simultaneously with a second RNA molecule while remaining bound to the NRE remains to be determined. If so, this would raise the possibility that the capacity to bring two different RNA molecules or two distant regions of the same RNA molecule into close proximity might be a widespread property of many such dual-RBD proteins whose RBDs are thought to function cooperatively in recognizing a shared RNA target.

In conclusion, the model that we have proposed for the RNA complex of the dual-RBDs of nucleolin plausibly accounts for much of our mutational data. This genetic approach constitutes a first step towards a high-resolution determination of the structure of this complex, which is currently being investigated by NMR and X-ray crystallography.

Materials and methods

Plasmids

The pRBD1+2 plasmid containing the two RBDs of nucleolin necessary and sufficient to confer the RNA-binding specificity of full-length protein (Serin *et al.*, 1997) was constructed by insertion of the RBD 1+2 gene as a 502 nucleotide *Nde*I–*Sal*I PCR fragment in the corresponding sites of pREV1 (Jain and Belasco, 1996), placing the RBD 1+2 synthesis under the control of an isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible promoter. Six additional amino acids (MRGSIH) are present at the N-terminus of the RBD 1+2 protein, but do not affect its binding affinity and specificity (data not shown). The parent plasmid is derived from pACYC184 and confers the resistance to chloramphenicol. The sequence of the RBD 1+2 protein is shown in Figure 1. pR1LL and pR2LL (with the mutation L43L45 and L125L127 in RBD 1 and 2 respectively) were constructed by insertion of the mutated RBD 1+2 cDNAs (as an *Nde*I–*Sal*I fragment) (Serin *et al.*, 1997) within the pRBD1+2 plasmid. To construct the pLac derivative plasmids, the sequence corresponding to the stem–loop IIB of the HIV-1 RRE was deleted from pLACZ-IIB plasmid (Jain and Belasco, 1996), and a *Bam*HI site was introduced one nucleotide upstream of the *lacZ* Shine–Dalgarno element to give pLacBHI plasmid. pNRE-*lacZ* was generated by introducing the oligonucleotide 5′ GATCATAAAGTGCAACCGAAATCCCGAAGTAGGAACAA 3′ hybridized to the complementary strand into the *Bam*HI site of pLacBHI. The underlined sequence corresponds to the 18 nucleotide motif identified by SELEX (Ghisolfi *et al.*, 1996) necessary and sufficient for a specific interaction of nucleolin. The construction of the different NRE mutants fused to the *lacZ* gene was performed following the same strategy, using oligonucleotides with the appropriate mutations. Plasmid constructions were confirmed by sequencing.

Random mutagenesis of the RBD 1+2 gene was performed using a PCR procedure (Cadwell and Joyce, 1992). Reaction conditions were: 0.5 mM $MnCl_2$, 7 mM $MgCl_2$, 10 mM Tris–HCl pH 8.4, 50 mM KCl, 30 pmol (each) of primers, 20 fmol of template, 1 mM dCTP and dTTP, 0.2 mM dATP and dGTP and 5 U of *Taq* polymerase (Boehringer) for 30 cycles of 94°C for 1 min, 45°C for 45 s, 72°C for 45 s. The mutagenized RBD 1+2 gene was then substituted for the corresponding fragment of pRBD1+2. A plasmid library (pRBD1+2M) of 25 000 independent clones was obtained by transforming the resulting ligation products into *E.coli*. Sequencing of 10 individual clones indicated the presence of one mutation every 300 nucleotides; therefore, on average, each RBD 1+2 gene (502 nucleotides) should contain 1–2 mutations.

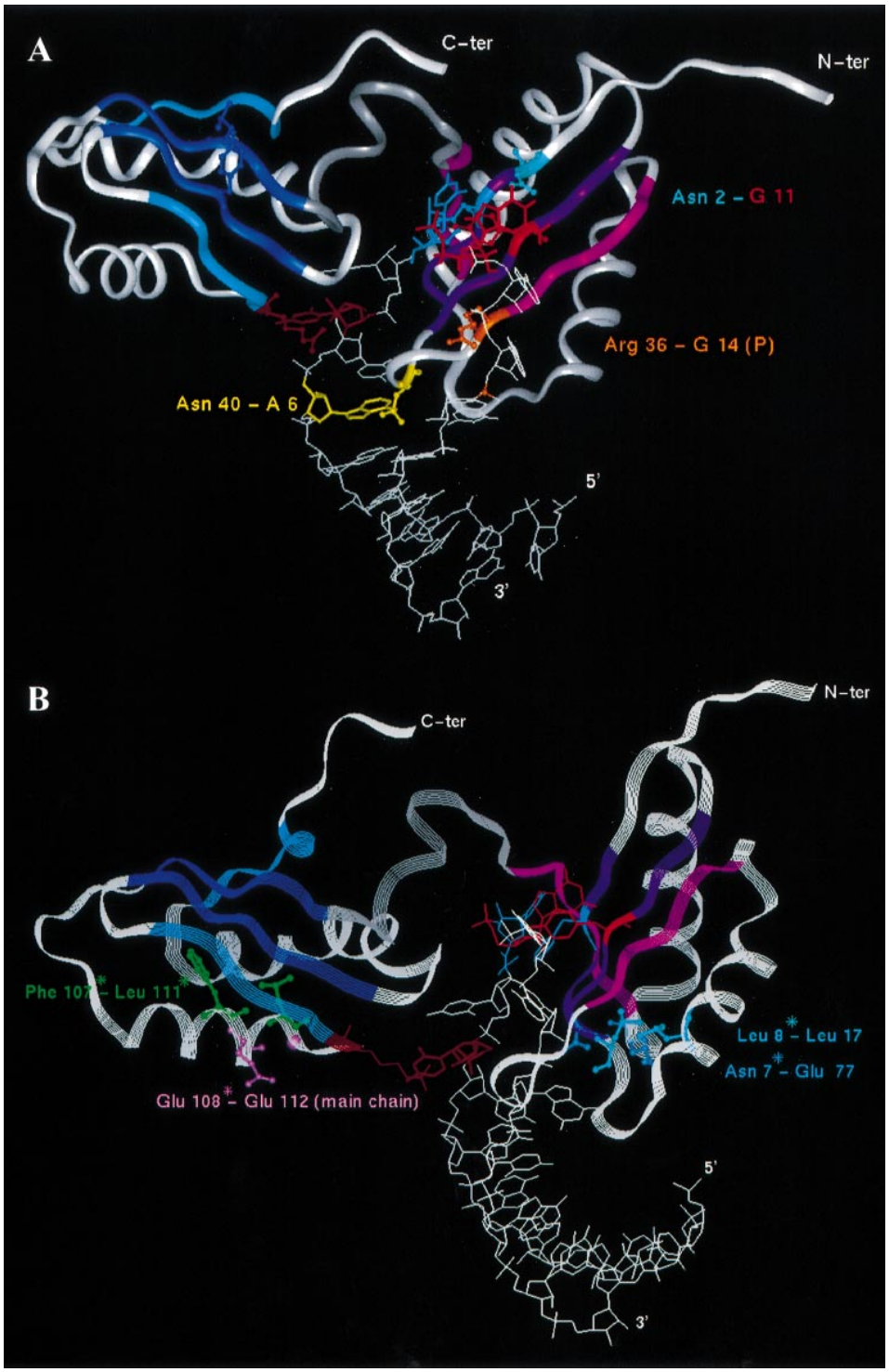

Fig. 5. (**A**) Proposed contacts between the R12 protein and the NRE. The color-coding of the contacts already described in Figure 4 is the same, blue for Phe4/C10, red for Tyr45/G11 and brown for Arg114/C8. Three new potential protein–RNA contacts are indicated. The critical positioning of G11 could be reinforced by a contact with Asn2 or Thr2 (in the case of the efficient NRE-binding N2T mutant displayed in Table IV). Arg36 could interact with the phosphate of G14, most likely through a hydrogen bond with one of its oxygens, and Asn40 amidic CO could form a hydrogen bond with A6 NH_2. Their respective mutants R36G and N40K do not mediate proper binding (see Table III). The side chains of the conserved RNP-2 and RNP-1 residues from RBD 2, Leu90 and Tyr127 respectively, are displayed in deep blue. (**B**) Key residues identified by the genetic screen and potentially involved in the stability of RBD 1+2 protein structure are shown. The asterisks designate those residues whose mutation has been shown to be detrimental to NRE binding. The amino acids with which they potentially interact have been displayed in the same color. Phe107 could form a hydrophobic interaction with Leu111, both residues being essential (see Table III). Glu108 acidic CO could form a hydrogen bond with Glu112 main-chain NH. These two pairs of interacting residues are in the close vicinity of Arg114, probably contributing to its critical positioning. Two other pairs of potential interacting residues are located in RBD 1, Leu8–Leu17 through hydrophobic contact, and Asn7–Glu77 through hydrogen bonding between amidic NH_2 and acidic CO. They are likely to stabilize the orientation of the RNP-2 motif (β1 strand) and hence the proper register of Phe4 with C10.

pR114Lib was generated by PCR mutagenesis with oligonucleotides 5′ GAGATCNNNTTGGTTAGCCCAGGATGGG 3′ and 5′ CTGGCTA-ACCAANNNGATCTCCAAGGC 3′, where the position of amino acid 114 was completely randomized.

Screening for altered-specificity RBD 1+2 variants

Preliminary experiments indicated that a high level of RBD 1+2 expression was toxic for *E.coli* cells. We therefore used an *E.coli* strain, WM1/F′ (*recA56 arg*$^-$ *lac-proXIII nal*^{r} *rif*^{r}*/F′lacI*^{q}), that makes high levels of the lac repressor and therefore significantly represses the expression of the RBD 1+2 protein in the absence of IPTG. This strain was used for all the experiments. Reporter plasmids and pRBD1+2 or its derivative plasmids were co-transformed into WM1/F′ strains using the CCMB procedure (Hanahan *et al.*, 1991) and plated on minimal medium containing glucose (0.4%), arginine (0.001%), proline (0.001%), 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, ampicillin (100 μg/ml), chloramphenicol (33 μg/ml) and X-gal (30 μg/ml), with or without 1 mM IPTG. Screening of altered specificity RBD 1+2 protein was then performed as described by Jain and Belasco (1996). Briefly, to identify mutations that disrupt the interaction of RBD 1+2 with wild-type NRE, pRBD1+2 and pLac-NRE were co-transformed in WM1/F′ and plated on the minimal plates containing X-gal, IPTG, ampicillin and chloramphenicol. After 36 h, blue colonies were picked up, and inoculated into LB medium with ampicillin, chloramphenicol and 1 mM IPTG until an OD_{600} of 0.3 was reached. Then, cultures were assayed for β-galactosidase activity and the production of full-length RBD 1+2 protein was checked by Western blot analysis. Colonies that express the RBD 1+2 protein variant to a level comparable with the wild type, and that were unable to repress the expression of NRE-*lacZ*, were selected and the corresponding plasmid pRBD1+2 mutant isolated for sequencing of the RBD 1+2 gene. Screening for suppressor mutations was performed with the same strategy, except that, in this case, colonies that were lighter blue in color than colonies with the wild-type pRBD1+2 plasmid were selected. Colonies were first re-selected on minimal plates with or without IPTG, then assayed for β-galactosidase activity and the production of full-length RBD 1+2 protein before isolation of the corresponding plasmids for sequencing.

β-Galactosidase assay in solution was performed as described by Miller (1972). Colonies were incubated overnight at 30°C in 2 ml of Luria–Bertani medium with ampicillin, chloramphenicol and 1 mM IPTG. Then, 2 ml of new medium, with IPTG, was inoculated with 60 μl of the overnight culture and incubated for 2 h at 37°C prior to harvesting and assaying for β-galactosidase activity. β-Galactosidase activities are given in Miller units: $OD_{420}\times10^3$/min/vol (ml)/OD_{600} and were determined for each reporter plasmid from at least two different transformations and with three independent colonies each time.

Protein production and purification

The wild-type RBD 1+2 gene and interesting variants were subcloned between the *Nde*I and *Bam*HI site of the pet15b plasmid (Novagen). Recombinant plasmids were transformed into the *E.coli* strain BL21-(DE3)plysS. Cells were grown at 37°C in LB (supplemented with 100 μg/ml ampicillin and 33 μg/ml chloramphenicol) until an OD_{600} of 0.5 was reached. Cells were induced with IPTG (1 mM) for 2 h, then rifampicin (150 μg/ml) was added, and the cultures were grown for a further 3 h at 37°C. Harvested cells were resuspended in buffer A (50 mM Na-phosphate pH 8, 300 mM NaCl) with DNase I (5 μg/ml) and lysed by sonication. After centrifugation (30 min at 10 000 *g*), the supernatant was recovered and gently mixed with 0.5 μl of Ni^{2+}-NTA resin (Qiagen) per ml of initial culture for 1 h at 4°C. After four washes with buffer A and four with buffer B (Na-phosphate 50 mM pH 6, 300 mM NaCl, 10% glycerol), tagged protein was eluted with buffer C (buffer B + 0.5 M imidazole). The supernatant was applied on a G-25 column (NAP 5-Pharmacia) equilibrated with 100 mM KCl and 10 mM Tris–HCl pH 7.5. Concentrations were estimated with Bradford reagent (Biorad protein assay) and checked by SDS–PAGE.

RNA transcription and RNA-binding assay

The pLac-NRE and pLac-NRE mutant plasmids were digested with *Taq*I. Transcription from the T7 polymerase promoter gives rise to a 64 nucleotide long RNA which contains the wild-type or mutated nucleolin recognition sequence flanked at its 3′ end by the Shine–Dalgarno sequence and the first four codons of the *lacZ* gene (see also Figure 1). RNAs were radiolabeled by incorporation of [α-^{32}P]CTP during transcription, according to the instructions of the manufacturer (Promega). RNAs were purified by repeated ammonium acetate (3 M) precipitation and their integrity checked by electrophoresis on a 6% acrylamide–8 M urea gel. [α-^{32}P]CTP incorporation was quantified to estimate RNA concentration.

For gel retardation assays, 10 fmol of labeled RNA were incubated in 10 μl of TMKC buffer (20 mM Tris–HCl pH 7.4, 4 mM $MgCl_2$, 200 mM KCl, 20% glycerol, 1 mM dithiothreitol, 0.5 mg/ml tRNA, 4 μg/ml bovine serum albumin) with the indicated amount of protein for 15 min at room temperature. The mixture was loaded directly on an 8% polyacrylamide gel (acrylamide:bis = 60:1) containing 5% glycerol in 0.5× TBE at room temperature. After electrophoresis, the gel was dried and subjected to autoradiography.

Molecular modeling of the RBD 1+2–NRE complex

Models were generated using the Biosym Technologies modules INSIGHTII, BIOPOLYMER, DISCOVER, DOCKING and HOMOLOGY (version 230), run on a Silicon Graphics Indigo Elan workstation.

The model of the 18 base stem–loop fragment corresponding to the NRE RNA was built in two steps. The stem was first built using the A-form RNA duplex parameters provided by the BIOPOLYMER module. Taking into account similarities between the U1 hairpin II and NRE loops (Ghisolfi *et al.*, 1996), we then used the atomic coordinates of the sugar–phosphate backbone of U1 hairpin II loop as a framework to build the NRE loop (residues P6–P12; accession number in the Brookhaven Protein Databank: pdb1urn). After appropriate base substitution and addition, the NRE stem and loop were merged and the resulting structure was refined by an energy minimization procedure consisting of 500 steps of steepest descent followed by 1000 cycles of conjugate gradients to a maximum derivative of 0.1 kcal/A/mol, using the DISCOVER consistent valence forcefield (CVFF).

Homology modeling of each nucleolin RBD was performed using a standard protocole (Greer, 1991) in the HOMOLOGY module and X-ray-derived coordinates of U1A first RBD as a reference (PDB code: pdb1urn). The assignment of the structurally conserved regions (SCRs) was based on the original RBD's alignment (Kenan *et al.*, 1991). The main modeling steps involved the transfer of coordinates between SCRs, the building of loops and a final structural refinement by energy minimization and molecular dynamics.

The 12 residue linker between the two nucleolin RBDs was first appended to RBD 1 in an extended structure and the peptidic bond made with the second RBD. Then appropriate Φ and ψ angles were assigned to each linker residue according to its predicted secondary structure (using both the Chou–Fasman and GOR II algorithms), a type-I′ turn structure for the four N-terminal amino acids (KGRD) (Smith and Pease, 1980) and a right-handed α-helical structure for the following five (SKKVR); several possible predicted configurations were explored for the last three residues (AAR), giving rise to a few potential RBD 1+2 structures. We kept as the most likely structure the one which did not present any steric clash and offered the quickest convergence when subjected to an extensive cycle of minimization and dynamics.

Docking of RBD 1+2 with the NRE stem–loop was performed within the DOCKING module by systematic searching for orientations of the RNA that offered a proper stacking interaction between RBD 1+2 Phe4 and Tyr45 and two NRE nucleotides. The best fit corresponding to the complex with the minimum energy of interaction was selected.

Acknowledgements

We thank D.Villa for help with the art work. This work was supported in part by grants from the Région Midi-Pyrénées (to M.E.) and a grant (NP-947) and a Faculty Research Award (FRA-419) to J.G.B. from the American Cancer Society.

References

Abdul-Manan,N., O'Malley,S.M. and Williams,K.R. (1996) Origins of the binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry*, **35**, 3545–3554.

Abe,R., Sakashita,E., Yamamoto,K. and Sakamoto,H. (1996) Two different RNA binding activities for the AU-rich element and the poly(A) sequence of the mouse neuronal protein mHuC. *Nucleic Acids Res.*, **24**, 4895–4901.

Allain,F.H.T., Gubser,C.C., Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature*, **380**, 646–650.

Avis,J.M., Allain,F.H.-T., Howe,P.W.A., Varani,G., Nagai,K. and Neuhaus,D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of the C-terminal residues in structure stability and RNA binding. *J. Mol. Biol*., **257**, 398–411.

Bentley,RC. and Keene,J.D. (1991) Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B′ protein and through interactions with U2 snRNP-A′ protein. *Mol. Cell. Biol*., **11**, 1829–1839.

Birney,E., Kumar,S. and Krainer,A.R. (1993) Analysis of the RNA recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res*., **21**, 5803–5816.

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

Burd,C.G. and Dreyfuss,G. (1994b) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J*., **13**, 1197–1204.

Caceres,J.F. and Krainer,A.R. (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J*., **12**, 4715–4726.

Cadwell,C.R. and Joyce,G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl*., **2**, 28–33.

Chung,S., Jiang,L., Cheng,S. and Furneaux,H. (1996) Purification and properties of HuD, a neuronal RNA-binding protein. *J. Biol. Chem*., **271**, 11518–11524.

Garret,D.S., Lodi,P.J., Shamoo,Y., Williams,K.R., Clore,G.M. and Gronenborn,A.M. (1994) Determination of the secondary structure and folding topology of an RNA binding domain of mammalian hnRNP A1 protein using three-dimentional heteronuclear magnetic resonance spectroscopy. *Biochemistry*, **33**, 2852–2858.

Greer,J. (1991) Comparative modeling of homologous proteins. *Methods Enzymol.*, **202**, 239–252.

Ghisolfi,L., Joseph,G., Puvion-Dutilleul,F., Amalric,F and Bouvet,P. (1996) Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA. *J. Mol. Biol*., **260**, 34–53.

Görlach,M., Wittekind,M., Beckman,R.A., Mueller,L. and Dreyfuss,G. (1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J*., **11**, 3289–3295.

Görlach,M., Burd,C.G. and Dreyfuss,G. (1994a) The determinants of RNA-binding specificity of the heterogenous nuclear ribonucleoprotein C proteins. *J. Biol. Chem*., **269**, 23074–23078.

Görlach,M., Burd,C.G. and Dreyfuss,G. (1994b) The mRNA poly(A) binding protein: localization, abundance, and RNA-binding specificity. *Exp. Cell Res*., **211**, 400–407.

Gubser,C.C. and Varani,G. (1996) Structure of the polyadenylation regulatory element of the U1A pre-mRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry*, **35**, 2253–2267.

Hanahan,D., Jesse,J. and Bloom,F.R. (1991) Plasmid transformation of *E.coli* and other bacteria. *Methods Enzymol*., **204**, 63–113.

Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.

Howe,P.W., Nagai,K., Neuhaus,D. and Varani,G. (1994) NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J*., **13**, 3873–3881.

Inoue,K., Hoshijima,K., Higuchi,I., Sakamoto,H. and Shimura,Y. (1992) Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. *Proc. Natl Acad. Sci. USA*, **89**, 8092–8096.

Jain,C. and Belasco,J.G. (1996) A structural model for the HIV-1 Rev–RRE complex deduced from altered-specificity Rev variants isolated by a rapid genetic strategy. *Cell*, **87**, 115–125.

Jessen,T.-H., Oubridge,C., Teo,C.H., Pritchard,C. and Nagai,K. (1991) Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J*., **10**, 3447–3457.

Jovine,L., Oubridge,C., Avis,J.M. and Nagai,K. (1996) Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy. *Structure*, **4**, 621–631.

Kanaar,R., Lee,A., Rudner,D.Z., Wemmer,D.E. and Rio,D.C. (1995) Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J*., **14**, 4530–4539.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci*., **16**, 214–220.

Kühn,U. and Pieler,T. (1996) *Xenopus* poly(A) binding protein: functional domains in RNA binding and protein–protein interaction. *J. Mol. Biol*., 256, 20–30.

Lapeyre,B., Bourbon,H. and Amalric,F. (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl Acad. Sci. USA*, **84**, 1472–1476.

LeCuyer,K.A., Behlen,L.S. and Uhlenbeck,O.C. (1996) Mutagenesis of a stacking contact in the MS2 coat protein–RNA complex. *EMBO J*., **15**, 6847–6853.

Lee,A.L., Kanaar,R., Rio,D.C. and Wemmer,D.E. (1994) Resonance assignments and solution structure of the second RNA-binding domain of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry*, **33**, 13775–13786.

Lutz-Freyermuth,C., Query,C.C. and Keene,J.D. (1990) Quantitative determination that one of two potential RNA-binding domains of a A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem–loop II of U1 RNA. *Proc. Natl Acad. Sci. USA*, **87**, 6393–6397.

Mayeda,A., Munroe,S.H., Caceres,J.F. and Krainer,A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J*., **13**, 5483–5495.

Merril,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogenous nuclear ribonucleoprotein. *J. Biol. Chem*., **263**, 3307–3313.

Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Nagai,K., Oudbridge,C., Ito,N., Avis,J. and Evans,P. (1995) The RNP domain: sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci*., **20**, 235–240.

Oubridge,C., Ito,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

Pakula,A.A., Young,V.B. and Sauer,R.T. (1986) Bacteriophage λ *cro* mutations: effects on activity and intracellular degradation. *Proc. Natl Acad. Sci. USA*, **83**, 8829–8833.

Sakashita,E. and Sakamoto,H. (1994) Characterization of RNA binding specificity of the *Drosophila* sex-lethal protein by *in vitro* ligand selection. *Nucleic Acids Res*., **22**, 4082–4086.

Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J*., **8**, 4163–4170.

Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B′ and their cognate RNAs. *Nature*, **345**, 502–506.

Serin,G., Joseph,G., Faucher,C., Ghisolfi,L., Bouche,G., Amalric,F. and Bouvet,P. (1996) Localization of nucleolin binding sites on human and mouse pre-ribosomal RNA. *Biochimie*, **78**, 530–538.

Serin,G., Joseph,G., Ghisolfi,L., Bauzan,M., Erard,M., Amalric,F. and Bouvet,P. (1997) Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J. Biol. Chem*., **272**, 13109–13116.

Shamoo,Y., Abdul-Manan,N., Patten,A.M., Crawford,J.K., Pelligrini, M.C. and Williams,K.R. (1994) Both RNA-binding domains in heterogeneous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry*, **33**, 8272–8281

Shamoo,Y., Abdul-Manan,N. and Williams,K.R. (1995) Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res*., **23**, 725–728.

Shamoo,Y., Krueger,U., Rice,L.M., Williams,K.R. and Steitz,T.A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol*., **4**, 215–222.

Smith,J.A. and Pease,L.G. (1980) Reverse turns in peptides and proteins. *CRC Crit. Rev. Biochem*., **8**, 315–399.

Srivastava,M., Fleming,P.J., Pollard,H.B. and Burns,A.L. (1989) Cloning and sequencing of the human nucleolin cDNA. *FEBS Lett*., **250**, 99–105.

Tacke,R. and Manley,J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J*., **14**, 3540–3551.

Tsai,D.E., Harper,D.S. and Keene,J.D. (1991) U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res*., **19**, 4931–4936.

Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.

Wittekind,M., Görlach,M., Friedrichs,M., Dreyfuss,G. and Mueller,L. (1992) ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry*, **31**, 6254–6265.

Xu,R.-M., Jokhan,L., Cheng,X., Mayeda,A. and Krainer,A.R. (1997) Crystal structure of human UP1, the two RNA-recognition motif domain of hnRNP A1. *Structure*, **5**, 559–570.

Zuo,P. and Manley,J.L. (1993) Functional domains of the human splicing factor ASF/SF2. *EMBO J*., **12**, 4727–4737.

Received on April 9, 1997; revised on June 2, 1997