Structural basis of the RNA-binding specificity of human U1A protein

Frédéric H.-T. Allain¹, Peter W.A. Howe²,
David Neuhaus and Gabriele Varani³

MRC Laboratory of Molecular Biology, Hills Road,
Cambridge CB2 2QH, UK

¹Present address: Kemisk Institut V, Copenhagen University,
Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

²Present address: Department of Chemistry and Biochemistry,
University of California, Los Angeles, CA 90095-1569, USA

³Corresponding author
e.mail: gv1@mrc-lmb.cam.ac.uk

The RNP domain is a very common eukaryotic protein domain involved in recognition of a wide range of RNA structures and sequences. Two structures of human U1A in complex with distinct RNA substrates have revealed important aspects of RNP–RNA recognition, but have also raised intriguing questions concerning the origin of binding specificity. The β -sheet of the domain provides an extensive RNA-binding platform for packing aromatic RNA bases and hydrophobic protein side chains. However, many interactions between functional groups on the single-stranded nucleotides and residues on the β -sheet surface are potentially common to RNP proteins with diverse specificity and therefore make only limited contribution to molecular discrimination. The refined structure of the U1A complex with the RNA polyadenylation inhibition element reported here clarifies the role of the RNP domain principal specificity determinants (the variable loops) in molecular recognition. The most variable region of RNP proteins, loop 3, plays a crucial role in defining the global geometry of the intermolecular interface. Electrostatic interactions with the RNA phosphodiester backbone involve protein side chains that are unique to U1A and are likely to be important for discrimination. This analysis provides a novel picture of RNA–protein recognition, much closer to our current understanding of protein–protein recognition than that of DNA–protein recognition.

Keywords: NMR/polyadenylation inhibition element/
RNA–protein interaction/RNP domain/U1A protein

Introduction

The RNP domain is one of the most common eukaryotic protein sequence motifs (Hodgkin, *et al.*, 1995), found in hundreds of RNA-binding proteins (Kenan *et al.*, 1991; Birney *et al.*, 1993; Burd and Dreyfuss, 1994a; Nagai *et al.*, 1995). Structural analysis of RNP domains from several different proteins (Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Wittekind *et al.*, 1992; Garrett *et al.*, 1994; Lu and Hall, 1995; Avis *et al.*, 1996) has demonstrated that the domain folds into a compact $\alpha\beta$ structure with a

four-stranded β -sheet packed against two α -helices through an extensive hydrophobic core. RNP proteins bind RNA with exquisite specificity utilizing this common structure. Site-directed mutagenesis (Nagai *et al.*, 1990; Jessen *et al.*, 1991) and NMR studies (Görlach *et al.*, 1992; Howe *et al.*, 1994; Kanaar *et al.*, 1995) identified the surface of the β -sheet as the primary site of RNA recognition. The crystal structure of the complex between the N-terminal domain of human U1A protein and stem-loop II of U1 snRNA (Oubridge *et al.*, 1994) revealed the structural basis for RNP–RNA recognition. The subsequent NMR structure of the same protein domain in complex with the polyadenylation inhibition element (Allain *et al.*, 1996), together with NMR structures of the free components (Avis *et al.*, 1996; Gubser and Varani, 1996), revealed that intermolecular recognition requires extensive conformational changes in both protein and RNA components.

The two structures of complexes of human U1A protein with distinct RNA substrates have clarified many important aspects of RNP–RNA recognition, but have also raised intriguing questions concerning the molecular basis of specificity. The structure of the U1A–hairpin complex (Oubridge *et al.*, 1994) showed that seven single-stranded nucleotides that represent the primary recognition site are involved in a very extensive network of hydrogen bonds with residues located on the surface of the β -sheet and in a loop immediately C-terminal to the domain. This observation would seem to imply that recognition of unpaired bases determines the specificity of RNP protein. This is somewhat puzzling, since the sequence diversity on the surface of the β -sheet is limited: RNP proteins are identified by highly conserved amino acids within the two central strands of the β -sheet itself (Birney *et al.*, 1993). How can this region form the basis for so many different and highly specific interactions of RNP proteins, when it is itself so highly conserved? Another puzzle is that two proteins (U1A and U2B^{''}) that share essentially all residues involved in recognition of the single-stranded nucleotides have different substrate specificities (Scherly *et al.*, 1990; Bentley and Keene, 1991).

The conservation of amino acids on the RNP domain β -sheet surface implies that much of its role is to provide a generic RNA-binding surface, while the specificity of the interaction is defined largely by the variable loops connecting the secondary structural elements of the protein (Kenan *et al.*, 1991). The refinement of the structure of the U1A complex with the polyadenylation inhibition element (PIE) reported here clarifies the role of the RNP domain specificity determinants (the variable loops of the domain) in recognizing a unique RNA three-dimensional structure. These results provide new insight into the molecular basis of the ability of RNP proteins to discriminate between RNA substrates.

Fig. 1. (A) Secondary structures of one of two repeated U1A-binding motifs from the polyadenylation inhibition element (left) and of stem-loop II of U1 snRNA (right); nucleotides involved in protein-RNA interactions are highlighted in bold. (B) Sequence and secondary structure of the N-terminal domain of human U1A protein.

Results

Structure refinement

Human U1A protein comprises two RNP domains separated by a linker containing a diffuse nuclear localization signal. The determinants of the affinity and specificity of this protein reside entirely within residues 1–102, i.e. the first RNP domain and the residues immediately C-terminal to it (Scherly *et al.*, 1990, 1991; Hall and Stump, 1992). The structure of the complex between a 30 nucleotide internal loop RNA derived from the U1A polyadenylation inhibition element (Figure 1A) and the U1A protein RNA-binding domain (amino acids 2–102) (Figure 1B) was determined using a very extensive set (~ 2600) of NMR-derived experimental constraints (Allain *et al.*, 1996). These data defined most of the structure and the intermolecular interface to high precision. However, critical parts of the interface (loop 1 and loop 3) were defined less precisely. The structure has been refined to improve the definition of these key parts of the intermolecular interface, resulting in the present precision of ~ 1 Å for the entire intermolecular interface (Table I).

The refinement was accomplished by using the previously reported set of experimentally determined structures (Allain *et al.*, 1996) to resolve ambiguous nuclear Overhauser effect (NOE) interactions. Novel intermolecular NOE contacts involve primarily resonances from the RNA sugars; these signals are poorly dispersed and overlap with protein H_α resonances. An illustration of the difficulties encountered during this process is presented in Figure 2; a detailed technical description of this process is presented elsewhere (Howe *et al.*, 1997). The new intermolecular NOE contacts involving H2' and H3' of A24 and H2' of G25 were particularly important to define interactions between A24 and G25 and loop 3 of U1A (residues 45–53). These intermolecular contacts, together

Table I. Statistics of NMR-derived experimental constraints and structural statistics for 31 converged structures

Distance and dihedral angle restraints	
NOE distances	
RNA–RNA	591
protein–protein	1710
RNA–protein	124
Hydrogen bonds	
protein	21
RNA	25
Dihedral angles (RNA)	110
Total no. of constraints	2581

Structure statistics

NOE violations	
number >0.2 Å	4 ± 1
maximum violation	<0.5 Å
Angle violations	
number $>5^\circ$	1.6 ± 1
Mean deviation from ideal covalent geometry	
bond length	0.05 Å
bond angles	0.80°
impropers	0.45°

R.m.s. deviations from mean structure (Å)

Protein	
all ordered residues (7–98) backbone	0.54 ± 0.07 Å
heavy atoms	0.98 ± 0.11 Å
RNA (all heavy atoms)	
all ordered regions	2.11 ± 0.72 Å
upper stem (G25–C38)	1.03 ± 0.37 Å
lower stem (G20–G23, C46–C49)	0.80 ± 0.11 Å
Protein-binding site (G23–G25, C38–C46)	1.01 ± 0.16 Å
single-stranded loop (A39–C45)	0.87 ± 0.11 Å
Protein–RNA (all heavy atoms)	
all ordered regions (7–98, G20–C49)	2.03 ± 0.61 Å
complex interface	1.08 ± 0.19 Å

with intramolecular interactions between A24 and G23, define precisely the relative position of the two helical stems beginning with G25–C38 and G23–C46. As a consequence of this improved definition, it has been possible to identify many more intermolecular interactions involving loop 3 and to establish that basic residues from loops 1 and 3 of U1A interact with the backbone phosphates of stem 2 (Figure 1A).

General features of the structure

A view of the refined structure is presented in Figure 3A. As previously reported (Allain *et al.*, 1996), the RNA is severely kinked in the complex. The RNA is recognized by amino acids located on the surface of the β -sheet and in the loops connecting the first strand of the β -sheet with helix A (loop 1), the second and third strands of the β -sheet (loop 3) and the loop connecting the end of $\beta 4$ to helix C. The intimate nature of the protein–RNA interface is particularly striking, the overall architecture of the complex being reminiscent of an enlarged protein structure. The RNA bases at the interface pack against hydrophobic protein side chains on the β -sheet surface to form what is in essence an enlarged hydrophobic core, while the negatively charged backbone phosphates of the RNA are on the surface of the structure, directed away from the protein and towards the solvent (Figure 3B). This arrangement is strikingly different from that of DNA–protein complexes. It also differs markedly from an early

Fig. 2. Sections of a three-dimensional ^{13}C -edited NOESY spectrum of the U1A-RNA spectrum containing intermolecular NOE interactions between Ser48 resonances and A24 sugar resonances.

model for RNP-RNA interaction for the related complex of U1A with stem-loop II of U1 snRNA (Jessen *et al.*, 1991; Howe *et al.*, 1994). In that model, the RNA bases were exposed to solvent while the phosphates interacted with the protein; the model explained the observation that mutations of nucleotides in the single-stranded loop cause only small decreases in affinity, while the phosphates are protected against chemical modifications (Jessen *et al.*, 1991).

Intermolecular interactions

A list of all intermolecular interactions observed in the refined structure is presented in Tables II–IV. Intermolecular contacts have been divided between hydrophobic interactions (van der Waals and stacking, Table II); hydrogen bonding (Table III); electrostatic interactions or salt bridges (Tables IV).

Intermolecular interactions involving the surface of the β -sheet and the seven common single-stranded nucleotides are very similar between the present structure and the crystal structure of the related hairpin complex (Figure 4). The root mean square deviation (r.m.s.d.) between the average of the present NMR ensemble and the crystal structure is 1.13 Å for the seven unpaired nucleotides and 1.29 Å for the portion of the protein-RNA interface involving those nucleotides. In this common region, differences between solution and crystal structures are most likely due to intrinsically greater flexibility in solution and to difficulties in detecting, by NMR, bound water molecules, which mediate several interactions in the crystal structure.

Novel interactions identified in this refined structure include electrostatic contacts from Arg47, Lys23 and Lys96 side chains to the RNA phosphates of stem 2 (Table IV and Figure 4B). These interactions, and others involving loop 3, are very different or absent in the hairpin complex (Figure 4C–D). In the present structure, the aromatic ring

of A24 approaches the protein backbone and interacts via two hydrogen bonds and hydrophobic interactions with Ser46. By contrast, U13, C14 and C15 in stem-loop II (which correspond roughly to A24 in the present structure) are poorly ordered in the hairpin complex and involved in crystal packing contacts (Oubridge *et al.*, 1994; Nagai *et al.*, 1995). The extensive contacts to G25 sugar and phosphates in the present complex have a single counterpart in the stem-loop 2 complex (an interaction involving G16 phosphate). Finally, Lys23, Ser46, Arg47 and Ser48 interact with the PIE RNA, but these residues are involved only in intramolecular interactions with other protein residues in the stem-loop II complex.

The differences between the two structures are directly supported by the primary data. Backbone amide resonances from loop 3 (Ser46–Arg52) are readily visible in ^1H – ^{15}N correlated spectra for the present complex, but were not detected in the corresponding spectrum recorded for the stem-loop II complex (data not shown). The absence of these signals from the spectrum of the stem-loop II complex probably reflects dynamic disorder of loop 3 in the solution conformation of the hairpin complex, due to less extensive intermolecular interactions. Therefore, U1A-RNA interactions are not identical between the two complexes, nor are they limited to the seven conserved nucleotides (5'AUUGCAC3') and the G-C base pair at the base of the loop.

In total, the NMR structure shows that residues from loop 3 (residues 46–52) form 18 hydrophobic contacts (Table II), eight hydrogen bonds (Table III) and two salt bridges with the RNA (Table IV) (Figure 4C), emphasizing the importance of this region of RNP proteins in RNA recognition. Val45 carbonyl and Arg47 and Leu49 main-chain amides make three hydrogen bonds with the RNA, while seven hydrophobic contacts involve the Ser46 and Ser48 side chains and one electrostatic interaction is observed from the Arg47 side chain. Additional interactions between loop 3 and the RNA involve the Ser48 hydroxyl and the side chains of Leu49, Lys50, Met51 and Arg52; these contacts were reported previously (Allain, *et al.*, 1996) and have been confirmed in this refined structure.

Discussion

Human U1A protein binds with very high affinity and specificity two distinct RNA targets; stem-loop II from U1 snRNA and the polyadenylation inhibition element (PIE) from the 3'-untranslated region of the U1A pre-mRNA (van Gelder *et al.*, 1993; Teunissen *et al.*, 1997). Binding to both RNAs occurs with a sub-nanomolar dissociation constant (Hall and Stump, 1992; van Gelder *et al.*, 1993; Hall, 1994; Gubser and Varani, 1996), while non-specific binding to poly(A), tRNA or double-stranded RNA is at least 10^6 -fold weaker (Hall and Stump, 1992). The numerous intermolecular interactions between U1A and either hairpin or internal loop targets could explain the very low dissociation constant and the large enthalpic contribution to binding (Hall and Kranz, 1995). However, the molecular basis for the ability of U1A and other RNP proteins to discriminate different RNAs remains an outstanding question.

Binding of U1A to hairpin and internal loop RNAs

Fig. 3. (A) Global view of a converged structure of the U1A-PIE complex. The RNA bases are splayed out across the surface of the β -sheet of U1A to form extensive intermolecular interactions, while the phosphates (dark blue) remain exposed to solvent. (B) The superposition of 31 converged structures highlights the excellent definition of the intermolecular interface. The backbone at the C-terminus of U1A N-terminal RNP domain is highlighted in yellow.

requires seven conserved single-stranded nucleotides and the presentation of these nucleotides within the correct secondary structure (van Gelder *et al.*, 1993; Hall, 1994). The single-stranded nucleotides are recognized primarily by residues from the protein β -sheet surface and from the loop connecting the end of the domain with helix C.

Intermolecular contacts involving these nucleotides are very similar between the two complexes of U1A (Oubridge *et al.*, 1994; Allain *et al.*, 1996), but interactions involving the protein variable loops differ significantly. In the hairpin complex (Oubridge *et al.*, 1994), U1A interacts only with the G-C base pair that closes the helical stem in addition

Table II. Hydrophobic and stacking interactions observed in the complex between human U1A protein and PIE RNA

Nucleotide	Atoms	Amino acid	Atoms
A24	C5, C6	Ser46	C α , C β
A24	C2', C3'	Ser48	C β
G25	C5'	Ser48	C β
G25	C8	Leu49	C α , C β
G25	C1', C4'	Leu49	C β
G25	C1'	Leu49	C δ 1
A39	C2	Leu49	C δ 1
U40	C1'	Leu49	C δ 2
G42	C2	Arg52	C β
C43	C2', C5, C6	Lys88	C ϵ
C43	C2, C4	Lys88	C γ
A44	C1', C4'	Met51	C ϵ
A44	C1'	Phe56	C ϵ , C ζ
A44	C2	Leu44	C δ
C45	C4	Ser90	C α
C45	C1'	Leu44	C δ 2
G42*	Base	Gln54	Side chain
C43*	Base	Tyr13	Side chain
A44*	Base	Phe56	Side chain
C45*	Base	Asn92	Side chain

Planar stacking interactions are indicated with *.

Table III. Hydrogen bonding interactions observed in the complex between human U1A protein and PIE RNA

Nucleotide	Atoms	Amino acid	Atoms
A24	N6-H	Val45	CO
A24	N7	Arg47	NH
G25	OP	Ser48	O γ -H
G25	O4'	Leu49	NH
G25	N7	Arg52	N η -H
A39	N1	Arg52	N η -H
U40	N3-H	Glu19	O ϵ
U41	N3-H	Asn16	O δ
G42	N1-H	Arg52	CO
G42	N2-H	Leu49	CO
G42	N7	Asn15	N δ 2-H
G42	O2'-H	Lys50	CO
C43	N3	Lys88	NH
C43	N4-H	Tyr86	CO
C43	O2	Lys88	NH
A44	N6-H	Thr89	CO
C45	N4-H	Asp90	CO
C45	O2	Ser91	O γ -H
C45	N3	Asp92	NH

Table IV. Electrostatic interactions observed in the complex between human U1A protein and PIE RNA

Nucleotide	Atoms	Amino acid	Atoms
A22	OP	Lys23	N ϵ
G23	OP	Lys23	N ϵ
G23	OP	Arg47	N η
C43	OP	Lys88	N ϵ
A44	OP	Lys50	N ϵ
C46	OP	Lys96	N ϵ

to the AUUGCAC sequence. An intact double-helical stem (which is essential for binding) was seen to restrict the conformational freedom of the single-stranded loop (thereby reducing entropic losses upon complex formation), and to allow recognition of the terminal G-C pair (Nagai *et al.*, 1995). The present structure contains more

extensive interactions between the RNA helical regions and amino acids from loop 1 and loop 3. Since these loops represent the principal specificity determinants of RNP proteins (Kenan *et al.*, 1991), the analysis of these interactions provides insight into how the specificity of U1A (and presumably other RNP proteins) is determined.

Interactions with the single-stranded nucleotides provide only limited ability to discriminate different RNA substrates

Nucleotides from the single-stranded A₃₉UUGCAC₄₅ sequence are splayed out against the protein β -sheet surface and the bases are involved in numerous intermolecular interactions. It is tempting to assume that recognition of the exposed functional groups on these bases suffices to define the specificity of U1A. As demonstrated by elegant *in vitro* selection experiments (Tsai *et al.*, 1991), these interactions are undoubtedly important determinants of U1A specificity. However, interactions with other regions of the RNA are also important.

The main consideration against a dominant role for residues from the β -sheet surface in determining RNP specificity is the extremely high sequence conservation in this region among different RNP proteins (Birney *et al.*, 1993). For example, protein residues involved in intermolecular stacking interactions (at positions 13, 54 and 56; Table II) correspond to either Phe or Tyr in $\sim 70\%$ of all RNP domains (Birney *et al.*, 1993). Although U1A has glutamine at position 54, this side chain also stacks (on G42), consistent with these intermolecular stacking interactions being general characteristics of RNP-RNA complexes. Similarly, the Ser46-A24 and Lys88-C43 hydrophobic interactions (Table II) can also be fulfilled by most of the equivalent residues in RNP proteins with diverse specificity. Thus (with the probable exception of Gln54; vide infra) these contacts are unlikely to be specific to U1A.

Nine intermolecular contacts involving direct side-chain interactions with base functional groups from the AUUGCAC sequence could be thought of as primary specificity determinants (Figure 4A). Seven of these contacts involve six residues on the surface of the β -sheet, while two involve residues of loop 3. Four of these residues, Asn15, Glu19, Leu44 and Gln85, are not well conserved in RNP proteins. These four residues contribute a total of six hydrogen bonds and are likely to be true specificity determinants on the surface of the β -sheet of U1A.

The present discussion provides an opportunity to re-examine the extensive mutagenesis data on the stem-loop II complex. Since interactions involving the single-stranded AUUGCAC sequence are common between the two structures, the effects of mutations here are likely to be very similar in the two cases. In fact, the C43 $\rightarrow$ U mutation within the PIE RNA (van Gelder *et al.*, 1993) and the corresponding change within U1 snRNA stem-loop II (Hall 1994) lead to an ~ 20 -fold increase in the dissociation constant in both cases.

Mutations of the conserved aromatic side chains on the β -sheet surface cause large reductions in affinity. Planar intermolecular stacking interactions are rare in DNA-protein complexes but have commonly been observed in RNA-protein complexes (Rould *et al.*, 1989, 1991;

Fig. 4. Intermolecular RNA-protein interactions in a low energy member of the family of converged structures. (A) Interactions between single-stranded nucleotide residues both in and immediately preceding helix C, and on the surface of the β -sheet; this view is identical to Figure 3B but a single structure is shown. (B) Electrostatic interactions involving stem 2 of the RNA and basic residues from loop 1 and loop 3. (C) Interactions of loop 3 in the present NMR structure and (D) in the crystal structure of the stem-loop II complex (Oubridge *et al.*, 1994).

Caverelli *et al.*, 1993; Biou *et al.*, 1994; Valegård *et al.*, 1994; Cusack *et al.*, 1996) and contribute significantly to binding energy (Stump and Hall, 1995; LeCuyer *et al.*, 1996). Mutation of Gln54 to Phe reduces the affinity >100-fold (Jessen *et al.*, 1991); this may reflect inability of the U1A-RNA interface to accommodate a larger side chain at this particular position. On the other hand, mutation of residues forming hydrogen bonding contacts from the amino acid side chains generally has a small effect (~ 10 -fold or less) on binding, with the exception of the Asn15 $\rightarrow$ Val substitution that abolishes binding altogether (Jessen *et al.*, 1991; Scherly *et al.*, 1991). Many mutations of single-stranded nucleotides also cause only small decreases (~ 10 - to 20-fold) in the binding constant (Hall 1994), even when residues involved in extensive intermolecular contacts are changed [for example C43 $\rightarrow$ U or C45 $\rightarrow$ G (Hall and Stump, 1992; Hall, 1994)]. The result of the C45 $\rightarrow$ G mutation is particularly surprising, since the structure does not allow any space for a purine in place of C45 and this nucleotide is important for discrimination between U1A and U2B' (Scherly, *et al.*, 1990). Only in two cases is the base identity truly critical: G42 $\rightarrow$ A and A44 $\rightarrow$ G transitions cause large decreases

($\sim 10^4$ -fold) in affinity (Hall, 1994). G42 is hydrogen bonded to two of the amino acid side chains that have been identified as primary determinants for U1A recognition of the single-stranded RNA loop, Asn16 and Glu19; however, Asn16 $\rightarrow$ Val and Glu19 $\rightarrow$ Asp mutations cause only small losses in binding constant (10-fold or less) (Scherly *et al.*, 1991). A44 forms a specific hydrophobic interaction with another unique amino acid side chain (Leu44).

An attractive explanation for the relative insensitivity of the binding constant to mutations of interfacial nucleotides is provided by considerations of flexibility. The small effect of the C45 $\rightarrow$ G substitution is inconsistent with a rigid interface, since there is insufficient space to fit a guanosine base in place of C45. Selective broadening of NMR resonances from residues close to C45 suggests that there are local motions in this region of the structure. Thus, the interface appears to be sufficiently flexible to accommodate either base through local conformational adjustments, as observed in crystallographic and thermodynamic studies of proteins (Alber *et al.*, 1987). Studies of residual dynamics at intermolecular protein-protein interfaces have led to the suggestion that a fine balance between rigidity and flexibility may provide a compromise

between complete specificity (at large entropic cost) and complete lack of selectivity (Kay *et al.*, 1996). The residual conformational flexibility may reduce the ability to discriminate between different nucleotides.

The analysis described here suggests that intermolecular interactions with the seven single-stranded nucleotides provide significant binding energy but only limited capacity to discriminate between different RNAs. Although the AUUGCAC sequence is optimal for U1A recognition (Tsai *et al.*, 1991), several contacts involving these nucleotides can be disrupted by mutations in the protein or RNA with only small effects on binding. Furthermore, many interactions involve either protein main-chain amide and carbonyl functionalities (Tables II–IV) or amino acids that are highly conserved among RNP proteins: these cannot be principal determinants of U1A specificity. To this extent, the surface of the β -sheet is indeed a generic RNA-binding surface (Kenan *et al.*, 1991), and the key molecular determinants of specificity must be sought in interactions involving loop 1, loop 3 and the region immediately C-terminal to the RNP domain.

The position of helix C is an important determinant of specificity

Sequences flanking the RNP domain are required for RNA binding in many RNP-containing proteins; often the isolated domain does not contain sufficient information to function as a sequence-specific RNA-binding entity (Kenan *et al.*, 1991). In the case of U1A, the region immediately following the end of the RNP domain is essential for RNA binding. A protein construct truncated at residue 91 does not bind RNA at all (Scherly *et al.*, 1989), truncation at residue 95 reduces binding 30-fold (Jessen *et al.*, 1991; Scherly *et al.*, 1991; Hall, 1994) and substitutions of Lys96 and Lys98 with Gln reduce affinity (Jessen *et al.*, 1991). Since sequence alignment suggests this region of the protein is unique to U1A (and U2B^o), interactions involving these residues are likely to contribute to the ability of U1A to discriminate different RNAs.

Residues 92–98 form a well-defined α -helix in free U1A (Howe *et al.*, 1994; Avis *et al.*, 1996), and this helical structure is preserved in the complex (Oubridge *et al.*, 1994; Allain *et al.*, 1996). Residues 88–92 are involved in extensive interactions with nucleotides C43, A44 and C45 (Figures 3B and 4A). Remarkably, almost all intermolecular contacts from this region of the protein involve main-chain functionalities. The reason why these interactions are nonetheless specific may lie in the conformational properties of helix C. The position of this helix in the complex is defined by hydrophobic interactions between Ile93, Ile94 and Met97 within helix C and His10, Leu41, Leu58 and Ile62. Mutations of some of the amino acids involved in positioning helix C (for example Thr11 next to His10) reduce RNA binding significantly (Jessen *et al.*, 1991). These residues and the residues within helix C are not preserved in other RNP proteins. Thus, these main-chain interactions become specific to U1A because the formation of helix C and its positioning in the complex through hydrophobic interactions with the rest of the domain are unlikely to occur in other RNP proteins.

A second mechanism by which helix C may contribute to molecular discrimination is related to its position in the

free protein. In the absence of RNA, helix C lies across the surface of the β -sheet and covers a large part of the RNA-binding surface (Figure 5B) (Avis *et al.*, 1996). Interactions between U1A and non-cognate RNAs may provide insufficient energy to drive this conformational change in the protein, thereby reducing the affinity for non-cognate RNAs.

Electrostatic interactions between the variable loops of the domain and the RNA phosphodiester backbone contribute to substrate discrimination

The variable loop 1 and loop 3 connecting the secondary structural elements of U1A (Figure 1B) contain an unusual cluster of basic residues (Lys20, Lys22, Lys23, Arg47, Lys50 and Arg52). The salt dependence of the binding constant for the hairpin complex suggests that electrostatic interactions contribute $\sim 30\%$ of the total binding energy (Hall and Stump, 1992; Hall, 1994). The role of these basic residues in U1A–PIE recognition is shown in Figure 4B, where electrostatic interactions between Lys23, Arg47 and Lys96 and the phosphodiester backbone of stem 2 are highlighted. Inspection of the electrostatic potential reveals that basic residues within U1A closely follow the path of the RNA phosphodiester backbone in both complexes (Nagai *et al.*, 1995; Allain *et al.*, 1996). One could argue that interactions with the phosphodiester backbone are of a non-specific nature. However, these interactions originate from residues in loop 1 and loop 3 that are unique to U1A (Birney *et al.*, 1993) and depend on RNA conformation, so they could well be specific for a particular RNA structure. The structure indicates that these interactions contribute to substrate discrimination by allowing the recognition of the charge distribution of the RNA substrate through electrostatic contacts. The importance of electrostatic interactions in RNP–RNA discrimination has been demonstrated for hnRNP A1, where an increase of the monovalent ion concentration reduces the ability to discriminate specific from non-specific substrates (Abdul-Manam *et al.*, 1996).

Loop 3–RNA interactions lock the conformation of the complex

Loop 3 (connecting strands $\beta 2$ and $\beta 3$) is the site of greatest diversity in sequence and length among RNP proteins. Residues from loop 3 of U1A form many intermolecular interactions with A24, G25–C38, A39 and U40 in the RNA at the junction between the single-stranded region and the two stems (Figure 4C). A particularly important role had been proposed for Arg52 (Oubridge *et al.*, 1994; Nagai *et al.*, 1995); in the stem–loop II complex, this residue recognizes the G25–C38 base pair and forms five hydrogen bonds. The hydrogen bonds involving Arg52 are defined less precisely in the present structure. Although Arg52 cannot be mutated to Gln, substitution of Arg52 by Lys causes only a 3-fold loss in binding constant (Nagai *et al.*, 1990; Jessen *et al.*, 1991). Taken together, these facts may be further pointers to conformational flexibility at the interface, as has also been observed in NMR structures of protein–DNA complexes (Berglund *et al.*, 1995; Slijper *et al.*, 1997).

The sequence requirements within loop 3 for binding to stem–loop II RNA have been investigated exhaustively in a genetic study based on the phage display method

Fig. 5. (A) The surface representation of the complex shows loop 3 of U1A (white) protruding through the hole in the RNA internal loop (light blue); interactions involving Leu49, Ser46 and Ser48 (dark blue) and Arg52 (hidden in this orientation) are critical to dock the protein against the RNA. (B) Surface representation of free and bound protein structures. The red surface identifies residues involved in intermolecular stacking interactions that become exposed upon the rearrangement of helix C; dark blue identifies loop 3 residues that interact with the RNA by rigid fit; light blue identifies the location of the remaining sites of intermolecular contact.

(Laird-Offinga and Belasco, 1995). Leu49 was conserved in all but one tight binding clone and was found to have a disproportionate effect on the kinetic parameters of binding (Leu49→Met substitution increased k_{off} 100-fold). In the present structure, the side chains of Leu49 and Arg52 interact with A39 and the G25-C38 base pair at the stem-loop junction. An identical interaction was observed for Leu136 with A72 stacking on the C71-G2 base pair in the acceptor stem of the tRNA^{Gln}-synthetase complex (Rould *et al.*, 1989). These hydrophobic interactions cannot simultaneously be fulfilled by any amino acid except leucine.

As shown in the surface representation of Figure 5A, contacts involving Ser46, Ser48 and Leu49 (and Arg52, not visible in this orientation) from loop 3 lock the protein into the hole defined by the RNA structure. Together with the stacking of A24 on stem 2 and A39 on stem 3, these interactions define the position of the helical stems with respect to the single-stranded nucleotides. In addition to intermolecular contacts involving the side chains of Ser46, Ser48, Leu49 and Arg52, every loop 3 residue except Ser46, Ser48 and Met51 forms intermolecular interactions involving main-chain functionalities. As in the case of the C-terminal region of the domain, these interactions may be specific because of the unique length (Birney *et al.*, 1993) and conformation of loop 3 in U1A, that forms a short helical structure in both free and bound U1A proteins.

Loop 3 residues are critical for discrimination between stem loop II of U1 snRNA and stem-loop IV of U2 snRNA (Scherly *et al.*, 1990; Bentley and Keene, 1991; Laird-Offinga and Belasco, 1995); the main difference between these RNAs is at the stem-loop junction. Disruption of the structure of the RNA stem-loop junction or an A39→C mutation (that would disrupt the interactions involving Leu49 and Arg52), reduces the affinity of U1A for its RNA by >1000-fold (Hall, 1994). Docking of loop 3 at the RNA stem-loop junction may be essential to define precisely the relative position of the double-helical stems (recognized through electrostatic interactions from basic residues in loops 1 and 3 to the phosphodiester backbone of stem 2) and single-stranded nucleotides (recognized by extensive interactions involving the protein β -sheet surface and helix C).

Recognition mechanism

An additional level of complexity in rationalizing RNP–RNA interaction concerns the extensive conformational changes observed in both protein and RNA components (Allain *et al.*, 1996; Avis *et al.*, 1996; Gubser and Varani, 1996). The major difference between the free and bound protein structures is in the position of helix C. In the complex, helix C is directed away from the β -sheet surface and the RNA (Figure 5B), while in the free structure it points towards loop 3 and covers part of the β -sheet surface (Avis *et al.*, 1996). Five of the seven single-stranded nucleotides (U41–C45) are highly flexible in the free RNA structure, but are much more highly ordered in the complex (Gubser and Varani, 1996). Distinct changes in base stacking interactions between the free and bound RNA structures are shown by the clear differences in the pattern of NOE interactions. Protein binding opens up the conformation of the RNA loop: in the free RNA, the A39, U40, U41, A44, C45 and A24 bases are oriented towards

the inside of the loop, filling the cavity created by the sugar–phosphate backbone of G23–G25 and C38–C46; only G42 and C43 are solvent exposed (Gubser and Varani, 1996). The space in the cavity occupied by U41, A44 and C45 in the free RNA is filled instead in the complex by protein residues from loop 3, while the bases of A24, A39, U40, U41, A44 and C45 are now directed towards the protein.

Implications for the binding specificity of other RNP proteins

The interconnection between RNA structure and specificity illustrated by the present structure provides a possible rationale to explain the high selectivity of RNP proteins that recognize highly structured RNAs, for example U1 70K. The numerous RNA–protein interactions mediated by highly conserved residues on the surface of the β -sheet are likely to be common to all RNP proteins, and may provide the free energy required for non-specific binding to any RNA target (Burd and Dreyfuss, 1994b). Substrate discrimination could be achieved by recognition of the unique three-dimensional shape and charge distribution of different RNAs, and fine-tuned by base-specific interactions with exposed single-stranded nucleotides. However, proteins that recognize genuinely single-stranded RNA must utilize a different recognition mechanism. Remarkably, single RNP domains fail to bind single-stranded RNA with high specificity: multiple domains are required for highly specific molecular recognition (Shamoo *et al.*, 1994; Kanaar *et al.*, 1995; Tacke and Manley, 1995; Tacke *et al.*, 1997). How this is achieved in structural terms remains unclear and poses the next logical question towards understanding RNP–RNA recognition.

Conclusions

The many structural and thermodynamic studies of DNA–protein complexes have provided a powerful paradigm in protein–nucleic acid recognition. Sequence-specific DNA binding requires recognition by protein side chains of functionalities on the DNA bases exposed in the major groove of B-form DNA (Steitz, 1990). This paradigm is so effective that a code, albeit a highly degenerate one, has been proposed for DNA recognition by zinc finger proteins (Choo and Klug, 1997). The implications of the structures of U1A–RNA complexes provide a very different picture of RNA–protein recognition, much closer to our understanding of protein–protein recognition than to DNA–protein recognition.

The recognition by U1A of the identity of RNA base functionalities exposed in single-stranded loops is an important element of molecular discrimination. However, many interactions with functional groups on the single-stranded nucleotides originate from the β -sheet surface and are potentially common to RNP proteins with diverse specificity. A large part of the role of the β -sheet is to form a generic RNA-binding surface by providing the amino acid side chains that participate in what is essentially an ‘intermolecular core’. The formation of intermolecular stacking interactions (Oubridge *et al.*, 1994; Allain *et al.*, 1996) through reorganization of the RNA and protein structures is an essential step in this process. Formation of a specific complex produces a buried interface that resembles the interior of a protein (Figure 3B). As observed

in studies of protein stability, the intermolecular interface may readily adjust to mitigate the effects of potentially deleterious substitutions (Alber *et al.*, 1987). Thus, although the sequence of the single-stranded nucleotides provides optimal intermolecular contacts and is therefore well conserved, mutations are surprisingly well tolerated. In contrast, inclusion of bulky amino acid side chains, disruption of Watson–Crick pairing in the helical stems or changes in the size of the single-stranded loop (Tsai *et al.*, 1991; Hall, 1994) reduce binding to nearly non-specific levels. The most variable region of RNP proteins, loop 3, plays a crucial role in defining the geometry of the intermolecular interface. Long-range electrostatic interactions with the phosphodiester backbone involve side chains that are unique to U1A and therefore likely to be important for discrimination.

Both protein and RNA structures acquire the correct conformation upon extensive conformational changes. This phenomenon has been observed for all RNA complexes with proteins and peptides for which structural data exist for both free and bound components (Puglisi *et al.*, 1992; Aboul-ela *et al.*, 1995; Peterson and Feigon, 1996). RNA–protein recognition is clearly a dynamic event: RNA structure defines distinct protein-binding surfaces, which are then reorganized upon interaction to optimize surface complementarity and functional group recognition. This flexibility of RNA structure has been exploited by proteins that regulate the activity of RNA enzymes (Weeks and Cech, 1995, 1996; Caprara *et al.*, 1996). An intriguing question yet to be addressed is to what extent these conformational changes are sequence-sensitive, in which case the ability to undergo such structural shifts could be as finely tuned and highly evolved as the RNA and protein structures themselves.

Materials and methods

All NMR experiments were conducted at 27°C in buffer containing either 10 mM sodium phosphate (pH ~6.5) or 5 mM sodium acetate (pH ~6). A detailed technical description of spectroscopic methods and the procedures adopted to obtain complete spectral assignments, to construct the constraint list and to calculate the structure of the complex is presented elsewhere (Howe *et al.*, 1997). Statistics for the constraints used in the calculations and all relevant structural statistics are reported in Table I.

The refinement of the structure presented here utilized the previously published structure (Allain *et al.*, 1996) to identify novel and ambiguous intermolecular NOE interactions. Fifty new intermolecular and 49 new intramolecular NOE constraints were identified during the refinement process. The majority of the novel intermolecular distance constraints (32 out of 49) involved RNA sugar resonances; given the extensive overlap in this region of the spectrum, these NOE interactions could only be identified conclusively using the first set of structures to resolve ambiguities. In addition, some constraints were removed from the original list because the previous assignments could not be validated conclusively in this more complete spectral analysis. The refinement of the constraint list resulted in a 20% improvement in the precision of the structure. A comparison between the two ensembles of structures revealed that the refined structures are within the envelope of structures defined by the first, less complete set of constraints. The 20% improvement in precision may seem disproportionate, since <100 novel constraints were added to a total of 2600 constraints. However, the majority of the novel constraints occur at the RNA–protein interface and often belong to the (previously) less well-defined regions of the structure.

Converged and non-converged structures were classified using clear differences in energy-ordered profiles and energy-ordered r.m.s.d. profiles (Avis *et al.*, 1996; Fletcher *et al.*, 1996; Varani *et al.*, 1996). In addition to improving the overall precision, the refined constraint set improved

significantly the proportion of structure calculations that converged and allowed a much clearer separation between converged and non-converged structures (Howe *et al.*, 1997). Thirty-one out of 50 structures have comparable low numbers of violations and low values for the pseudo-energy corresponding to NOE distance constraints; violations increase significantly for structures 32–50, and the energy-ordered r.m.s.d. profile diverges from the clear plateau established for the first 31 structures (Howe *et al.*, 1997). All statistics reported in Table I and the analysis of intermolecular interactions were based on the ensemble of 31 converged structures.

Intermolecular interactions (Tables II–IV) have been defined by a statistical analysis based on a systematic search through all converged structures for hydrogen bonding, van der Waals and electrostatic contacts. Interactions were considered to be established only when observed in the majority of all converged structures. Since NMR structures are constructed from numerous, short interproton distances, stacking, hydrophobic and hydrogen bonding contacts are defined much more precisely than electrostatic interactions or salt bridges. Furthermore, we could not observe any direct NOE interaction between basic protein side chains and the RNA phosphodiester backbone. Therefore, the proposed electrostatic contacts could only be inferred indirectly by a statistical analysis of converged structures. We emphasize, nevertheless, that the electrostatic component of the force field was never introduced during any stage of the structure calculation process, thereby avoiding any bias that would favour interactions between basic amino acids and the RNA phosphates.

Acknowledgements

We would like to thank Drs Johanna Avis, Charles Gubser, Kiyoshi Nagai and Chris Oubridge for help and suggestions at various stages of this project.

References

- Abdul-Manam, N., O'Malley, S.M. and Williams, K.R. (1996) Origins of the binding specificity of the A1 heterogeneous ribonucleoprotein. *Biochemistry*, **35**, 3545–3554.
- Aboul-ela, F., Karn, J. and Varani, G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.*, **253**, 313–332.
- Alber, T., Dao-Pin, S., Wilson, K., Wozniak, J.A., Cook, S.P. and Matthews, B.W. (1987) Contributions of hydrogen bonds of Thr157 to the thermodynamic stability of phage T4 lysozyme. *Nature*, **330**, 41–46.
- 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., Allain, F.H.-T., Howe, P.W.A., Varani, G., Neuhaus, D. and Nagai, K. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
- Bentley, R.C. 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.
- Berglund, H., Baumann, H., Knapp, S., Ladenstein, R. and Härd, T. (1995) Flexibility of an arginine side chain at a DNA–protein interface. *J. Am. Chem. Soc.*, **117**, 12883–12884.
- Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA^{Ser}. *Science*, **263**, 1404–1410.
- 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.
- Caprara, M.G., Mohr, G. and Lambowitz, A.M. (1996) A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J. Mol. Biol.*, **257**, 512–531.

- Caverelli, J., Rees, B., Ruff, M., Thierry, J.-C. and Moras, D. (1993) Yeast tRNA^{Asp} recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature*, **362**, 181–184.
- Choo, Y. and Klug, A. (1997) Physical basis of a protein–DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.
- Cusack, S., Yaremchuk, A. and Tuko, M. (1996) The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with tRNA^{Ser} and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.*, **15**, 2834–2842.
- Fletcher, C.M., Jones, D.N.M., Diamond, R. and Neuhaus, D. (1996) Treatment of NOE constraints involving equivalent or nonstereoisotopic protons in calculations of biomacromolecular structures. *J. Biomol. NMR*, **8**, 292–310.
- Garrett, 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-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry*, **33**, 2852–2858.
- 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.
- Gubser, C.C. and Varani, G. (1996) Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry*, **35**, 2253–2267.
- Hall, K.B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*, **33**, 10076–10088.
- Hall, K.B. and Kranz, J.K. (1995) Thermodynamics and mutations in RNA–protein interactions. *Methods Enzymol.*, **259**, 261–281.
- Hall, K.B. and Stump, W.T. (1992) Interaction of N-terminal domain of U1A protein with an RNA stem-loop. *Nucleic Acids Res.*, **20**, 4283–4290.
- Hodgkin, J., Plaster, R.H.A. and Waterston, R.H. (1995) The nematode *C. elegans* and its genome. *Science*, **270**, 410–414.
- 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.A., 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.
- Howe, P.W.A., Allain, F.H.-T., Varani, G. and Neuhaus, D. (1997) Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. Biomol. NMR*, in press.
- Jessen, T.H., Oubridge, C., Teo, C.H., Pritchard, C. and Nagai, K. (1991) Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.*, **10**, 3447–3456.
- Kanaar, R., Lee, A.L., 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.
- Kay, L.E., Muhandiram, D.R., Farrow, N.A., Aubin, Y. and Forman-Kay, J.D. (1996) Correlation between dynamics and high affinity binding in an Sh2 domain interaction. *Biochemistry*, **35**, 361–368.
- Kenan, D.J., Query, C.C. and Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.
- Laird-Offing, I.A. and Belasco, J.G. (1995) Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc. Natl Acad. Sci. USA*, **92**, 11859–11863.
- 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.
- Lu, J. and Hall, K.B. (1995) An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol.*, **247**, 739–752.
- Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Nagai, K., Oubridge, C., Ito, N., Avis, J. and Evans, P. (1995) The RNP domain: a 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.
- Peterson, R.D. and Feigon, J. (1996) Structural change in Rev responsive element RNA of HIV-1 on binding Rev peptide. *J. Mol. Biol.*, **264**, 863–877.
- Puglisi, J.D., Tan, R., Canlan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR–arginine complex by NMR. *Science*, **257**, 76–80.
- Rould, M.A., Perona, J.J., Söll, D. and Steitz, T.A. (1989) Structure of *E. coli* glutamyl tRNA-synthetase complexes with tRNA^{Gln} and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
- Rould, M.A., Perona, J.J. and Steitz, T.A. (1991) Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. *Nature*, **352**, 213–218.
- 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. (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B' and their cognate RNAs. *Nature*, **345**, 502–506.
- Scherly, D., Kambach, C., Boelens, W., van Venrooij, W.J. and Mattaj, I.W. (1991) Conserved amino acid residues within and outside of the N-terminal ribonucleoprotein involved in U1 RNA binding. *J. Mol. Biol.*, **219**, 577–584.
- Shamoo, Y., Abdul-Manam, N., Patten, A.M., Crawford, J.K., Pellegrini, 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.
- Slijper, M., Boelens, R., Davis, A.L., Konings, R.N.H., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1997) Backbone and side chain dynamics of the *lac* repressor headpiece (1–56) and its complex with DNA. *Biochemistry*, **36**, 249–254.
- Steitz, T.A. (1990) Structural studies of protein–nucleic acid interactions: the sources of sequence-specific binding. *Q. Rev. Biophys.*, **23**, 205–280.
- Stump, W.T. and Hall, K.B. (1995) Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA*, **1**, 55–63.
- 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.
- Tacke, R., Chen, Y. and Manley, J.L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl Acad. Sci. USA*, **94**, 1148–1153.
- Teunissen, S.W.M., van Gelder, C.W.G. and van Venrooij, W.J. (1997) Probing the 3' UTR structure of U1A mRNA and footprinting analysis of its complex with U1A protein. *Biochemistry*, **36**, 1782–1789.
- 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.
- Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L. (1994) Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature*, **371**, 623–626.
- van Gelder, C.W.G., Gunderson, S.I., Jansen, E.J.R., Boelens, W.C., Polycarpou-Schwartz, M., Mattaj, I.W. and van Venrooij, W.J. (1993) A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.*, **12**, 5191–5200.
- Varani, G., Aboul-ela, F. and Allain, F.H.-T. (1996) NMR investigations of RNA structure. *Prog. NMR Spectrosc.*, **29**, 51–127.
- Weeks, K.M. and Cech, T.R. (1995) Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5' splice site domain. *Cell*, **82**, 221–230.
- Weeks, K.M. and Cech, T.R. (1996) Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science*, **271**, 345–348.
- Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. and Mueller, L. (1992) ¹H, ¹³C, and ¹⁵N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry*, **31**, 6254–6265.

Received on April 28, 1997; revised on July 2, 1997