Article No. jmbi.1999.3311 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 967–979

JMB

Changes in Side-chain and Backbone Dynamics Identify Determinants of Specificity in RNA Recognition by Human U1A Protein

Anthony Mittermaier[1], Luca Varani[2], D. R. Muhandiram[1], Lewis E. Kay[1] and Gabriele Varani[2]*

[1]*Protein Engineering Network Centers of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada*

[2]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom*

**Corresponding author*

The ribonucleoprotein (RNP) domain is one of the most common eukaryotic protein domains, and is found in many proteins involved in recognition of a wide variety of RNAs. Two structures of RNA complexes of human U1A protein have revealed important aspects of RNP-RNA recognition, but have also raised intriguing questions concerning how RNP domains discriminate between different RNAs. In this work, we extend the investigation of U1A-RNA recognition by comparing the dynamics of U1A protein both free and in complex with RNA. We have also investigated the trimolecular complex between two U1A proteins and the complete polyadenylation inhibition element to study the effect of RNA-dependent protein-protein interactions on protein conformational flexibility. We report that changes in backbone dynamics upon complex formation identify regions of the protein where conformational exchange processes are quenched in the RNA-bound conformation. Furthermore, amino acids whose side-chains experience significant changes in conformational flexibility coincide with residues particularly important for the specificity of the U1A protein/RNA interaction. This study adds a new dimension to the description of the coordinated changes in structure and dynamics that are critical to define the biological specificity of U1A and other RNP proteins.

Keywords: NMR; polyadenylation inhibition element; RNP domain; RNA-protein interaction; U1A protein

Introduction

The ribonucleoprotein (RNP) domain is one of the most common eukaryotic protein sequence motifs (The *C. elegans* Consortium, 1998). It has been identified in hundreds of RNA-binding proteins and has been implicated in the specific recognition of a wide range of RNA sequences and structures (Varani & Nagai, 1998). 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 NMR structure of the same protein domain in complex with the polyadenylation inhibition element (PIE) RNA (Allain *et al.*, 1996), together with the structures of the free components (Avis *et al.*, 1996; Gubser & Varani, 1996), revealed that intermolecular recognition requires extensive conformational changes in both protein and RNA components. Protein-induced RNA conformational changes or disorder-order transitions have been observed in essentially all protein-RNA complexes studied so far (Varani, 1997). The question of how binding and specificity are coupled to these conformational changes and to the dynamic processes that occur during these interactions is an important but still unaddressed issue in RNA-protein recognition. The information existing on the U1A system makes it a paradigm in RNP-RNA recognition, and therefore an ideal system to address these central questions.

A. Mittermaier & L. Varani contributed equally to this work.

Abbreviations used: RNP, ribonucleoprotein; snRNA, small nuclear RNA; PIE, polyadenylation inhibition element; NOE, nuclear Overhauser enhancement; HSQC, heteronuclear single quantum coherence.

E-mail address of the corresponding author: gv1@mrc-lmb.cam.ac.uk

0022-2836/99/490967–13 $30.00/0

All RNPs share the same topology and three-dimensional structure, a split αβ fold packing a four-stranded β-sheet against two α-helices on the surface opposite RNA binding (Varani & Nagai, 1998). Residues involved in RNA recognition cluster on the surface of the β-sheet, in the variable loops connecting the β-strands (loop 1 and 3) and in the regions that immediately follow or precede the domain (Allain *et al.*, 1996; Oubridge *et al.*, 1994). Many of these residues are highly conserved among all RNPs (Kenan *et al.*, 1991), raising the question of how discrimination is achieved. In order to understand better how RNP proteins discriminate between different RNAs, we have extended the study of U1A-RNA recognition through the comparison of backbone and side-chain dynamics of U1A protein both free and in complex with RNA. We show that residues that are particularly important for specificity display significant changes in backbone or side-chain dynamics upon binding to RNA. The further extension of the investigation to the cooperative trimolecular complex formed by two U1A proteins and the complete PIE RNA (Figure 1) reveals changes in protein dynamics that reflect the formation of RNA-dependent protein-protein interactions.

Results

Backbone dynamics

Backbone ^{15}N relaxation data and side-chain methyl-^{2}H relaxation data were recorded at 600 MHz on free U1A protein (residues 2-102 of human U1A), fully ^{15}N-^{13}C labeled and randomly fractionally deuterated to an extent of approxi-

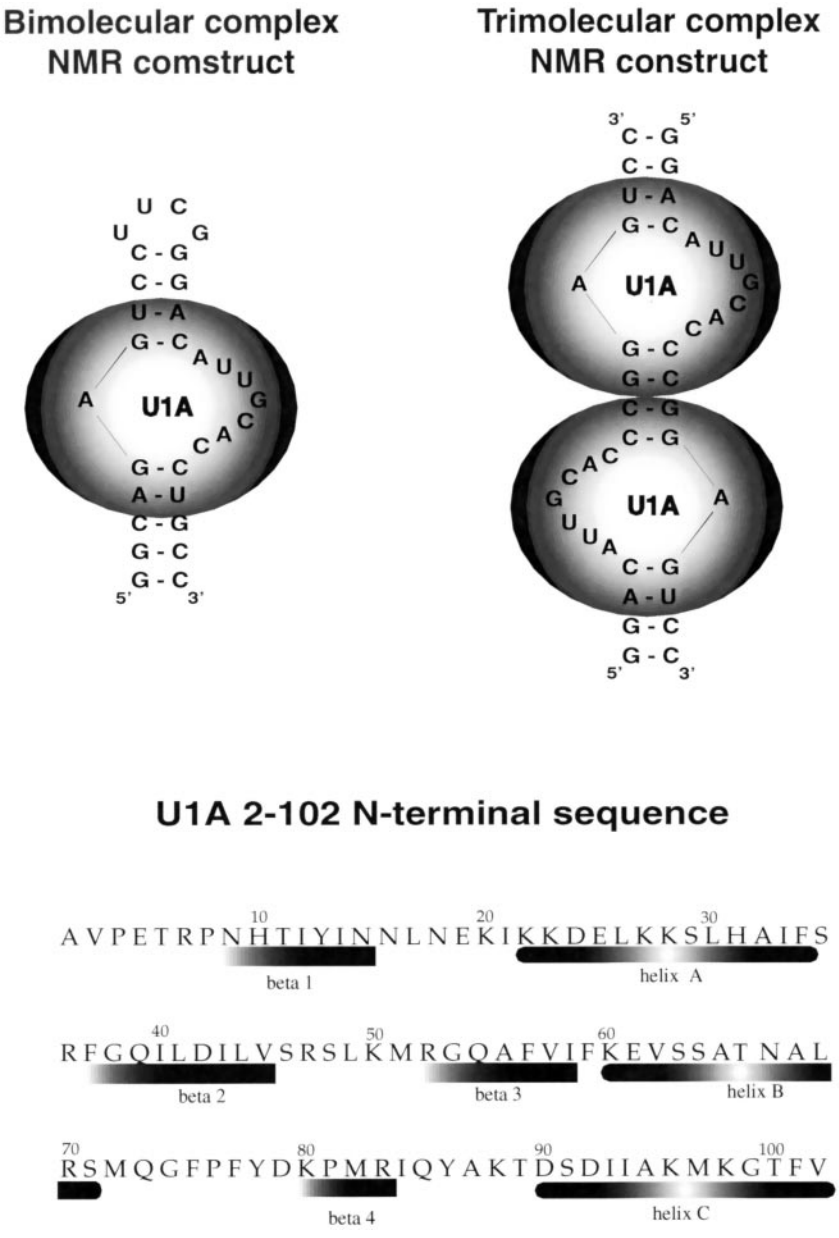

Figure 1. Sequence and secondary structure of the RNAs used in the study of the bimolecular (left) and trimolecular (right) complexes. U1A protein is shown schematically as shaded spheres; its sequence and secondary structure are shown in the bottom panel.

mately 50 %. Similar data were also recorded on the complex of U1A with one of the two high affinity sites within the U1A PIE. In addition, ^{15}N relaxation data were obtained for the trimolecular complex composed of two U1A molecules and the complete PIE RNA (Figure 1). PIE RNA contains two closely related protein-binding sites that differ at two nucleotides, but a symmetrical RNA is fully functional in inhibition of polyadenylation (Gunderson *et al.*, 1997). We studied this fully symmetrical complex (Figure 1) to reduce spectral complexity.

The ^{15}N T_1, T_2 and ^1H-^{15}N heteronuclear nuclear Overhauser enhancements (NOEs) were measured for free and bound U1A protein by standard methods (Farrow *et al.*, 1994), as reported in Methods. Heteronuclear correlated spectra were well resolved in the amide and methyl regions of ^1H-^{15}N and ^1H-^{13}C heteronuclear single quantum coherence (HSQC) spectra for both free and bound states of U1A. Representative ^{15}N T_2 relaxation decay curves for selected amino acids of the free protein and the bimolecular complex are shown in Figure 2(a) and (b). Although the increase in molecular mass leads to an increase in the uncertainty of the individual T_1 and T_2 measurements, the quality of the data remain very high in the complexes. Accurate and precise ^{15}N relaxation parameters could therefore be obtained for free U1A protein, as well as for the bimolecular and trimolecular complexes. The ^{15}N T_2 values and experimental uncertainties are compared in Figure 3 for free U1A protein and for the bimolecular and trimolecular U1A-RNA complexes.

The ^{15}N T_1 values are between 520 and 600 ms for all residues in the free U1A protein, with the exception of unfolded residues at the N and C termini with slightly longer T_1 values (600-720 ms)

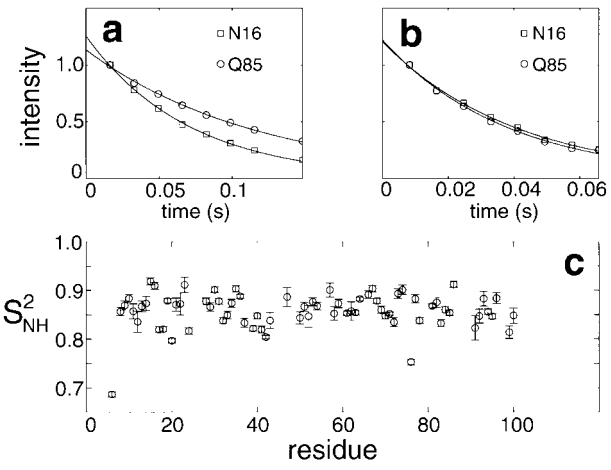

Figure 2. Normalized peak intensities as function of relaxation delay from the ^{15}N T_2 relaxation experiments recorded on (a) the free protein and (b) the bimolecular complex. (c) Values of the order parameters (S^2) for the free protein calculated assuming an axially symmetric diffusion tensor with $D_{iso} = 2.72 \times 10^7$ s^{-1} and $D_{par}/D_{perp} = 0.90$.

(data not shown). Very similar T_1 values were reported for a shorter construct, U1A (1-95) (Lu & Hall, 1997). Upon formation of the bimolecular complex, T_1 values approximately double to approximately 1.1-1.4 seconds, with the exception of residues at either end of the protein domain where T_1 values remain comparable to those observed for the same residues in the free protein. The termini of the domain remain poorly ordered upon formation of the protein-RNA complex (Allain *et al.*, 1996, 1997). A further increase in the average T_1 value is observed in the trimolecular complex, typical T_1 values being 1.6-1.8 seconds. ^{15}N T_1 values increase only slightly for residues Val3-Thr6 at the N terminus of the domain, but lengthen somewhat at the opposite end of the molecule. ^{15}N T_2 values are plotted in Figure 3(a) for free U1A protein. The values of 110-120 ms observed for the well-ordered regions of free U1A protein are very similar to those previously reported for a shorter construct (120 ms) (Lu & Hall, 1997). As expected, these values are longer in the disordered N-terminal tail of the domain.

The ^{15}N T_2 values are shorter than average for some residues in the two central strands of the β-sheet (β1 and β3), within loop 3 (connecting β2 and β3), within helix C and in the loop connecting β4 with helix C. Furthermore, several residues from these same regions appear broad in ^1H-^{15}N HSQC spectra of free U1A protein (residues 2-102). As discussed below, when the relaxation data from these residues are fit using the model free formalism (Lipari & Szabo, 1982), significant R_{ex} terms are needed (Figure 4). These observations suggest very strongly that conformational exchange processes occur in these regions of the protein which coincide with the RNA-binding interface (see below). The ^{15}N T_2 values decrease significantly in the bimolecular RNA complex and become much more uniformly distributed around an average value of 35-40 ms (Figure 3(b)). Only the N (Val3-Glu5) and C-terminal (Lys96-Phe101) residues maintain longer T_2 values, suggesting that these regions of the protein remain highly mobile. In contrast to what was observed for the free U1A protein, residues within loop 3 and at the hinge region preceding helix C (Thr89-Ser91) are clearly visible in spectra of the complex and have T_2 values comparable to those observed for the well-ordered amino acid residues in the protein domain. Thus, conformational exchange within loop 3 and the RNP-helix C hinge region is quenched upon RNA binding. The implications of this observation for RNA recognition are analyzed in the Discussion.

In the trimolecular complex formed by two U1A proteins and the complete PIE RNA (Figure 1),^{15}N T_2 values decrease further to 25-30 ms, as expected from the increase in molecular mass, with the exception of the N-terminal unfolded tail. The only significant difference between bimolecular and trimolecular complexes is found within the C-terminal end of the protein, which is significantly more

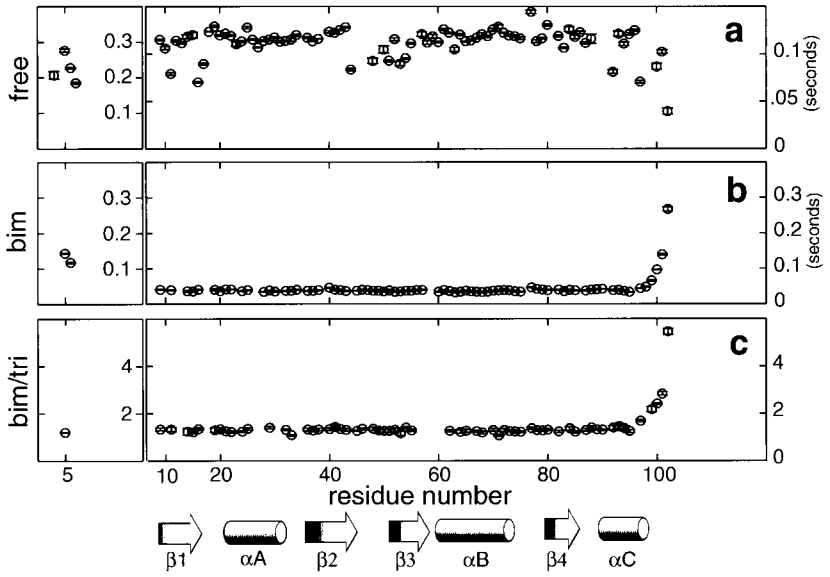

Figure 3. ^{15}N T_2 values of (a) free U1A, (b) the bimolecular complex of U1A and the half-site RNA. (c) The ratio of ^{15}N T_2 values measured for the bimolecular and trimolecular complexes (T_2, bimolecular/T_2, trimolecular). The protein secondary structure is schematically shown at the bottom of the Figure.

rigid in the trimolecular complex. This is best illustrated by plotting the ratio of T_2 values measured for the bimolecular and trimolecular complex. Relative to the rest of the domain, significantly larger T_2 ratios are observed for amino acids in this region of the protein (Figure 3(c)), reflecting much larger than average reductions in T_2 for these residues in the trimolecular complex. This behavior reflects the formation of protein-protein contacts in the trimolecular complex that extend helix C by a further turn and make it significantly more ordered on the time scales sampled by ^{15}N relaxation (see below).

Quantitative analysis of the ^{15}N relaxation data

An anisotropic diffusion tensor was used to interpret the ^{15}N T_1 and T_2 values for the free protein using the local diffusion constant approach (Bruschweiler *et al.*, 1995). An axial symmetric model of diffusion was found to afford a statistically significant reduction in χ^2 ($p = 0.005$) over a single rotational diffusion constant, while a fully asymmetric model did not significantly improve the fit. The value of D_{iso} ($D_{iso} = (D_{xx} + D_{yy} + D_{zz})/3$, where D is the diagonalized diffusion tensor) obtained from this analysis, corresponds to a τ_m of 6.1 ns. This is typical for a protein of approximately 100 residues and is close to the value reported for a shorter fragment of the same protein (Lu & Hall, 1997). The degree of anisotropy was found to be small, with $D_{par}/D_{perp} = 0.90$ ($D_{par} = D_{zz}$ and $D_{perp} = D_{xx} = D_{yy}$ for an axially symmetric diffusion tensor). The good agreement between experimental local diffusion constants and those predicted by an axially symmetric model of diffusion is shown in Figure 4(a). Residues experiencing chemical exchange have anomalously low values for the apparent local diffusion constant and fall beneath the line predicted by the model. From the axially symmetric diffusion parameters, ideal values of local τ_m were calculated for each NH bond vector according to (Lee *et al.*, 1997):

$$\tau_m = (6D_{iso} - 2P_2(\cos\theta)(D_{par} - D_{perp}))^{-1} \quad (1)$$

where θ is the angle between the NH bond vector and the principal (unique) axis of the diffusion tensor and $P_2(x) = (3x^2 - 1)/2$. Dynamics parameters were obtained on a per residue basis by minimization of a target function of the form:

$$\chi^2 = (T_{1c} - T_{1e})^2/\sigma_{T1}^2 + (T_{2c} - T_{2e})^2/\sigma_{T2}^2 + (NOE_c - NOE_e)^2/\sigma_{NOE}^2 \quad (2)$$

where the subscripts c and e represent calculated and experimentally determined relaxation parameters, respectively. Standard equations for T_1, T_2 and NOE were used (Kay *et al.*, 1989), and the form of the spectral density function was assumed to be (Lipari & Szabo, 1982):

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2) \quad (3)$$

with $1/\tau = 1/\tau_m + 1/\tau_e$. In equation (3), S^2 is the order parameter describing the amplitude of pico-nanosecond time-scale motion, τ_e and τ_m are the correlation times describing internal motion and molecular tumbling, respectively. Order parameters obtained in this analysis are shown in Figure 2(c), and are between 0.8 and 0.9 for the well-folded regions of the free protein, with the single exception of Phe77. These values are consistent with order parameters generally found in studies of folded protein domains. Results obtained for R_{ex} are shown in Figure 4(b). Large values (1 to 6 s^{-1}) were found for some residues in strands β1 and β2, loop3 and helix C.

We attempted to analyze the relaxation data for the RNA-protein complexes as well using the

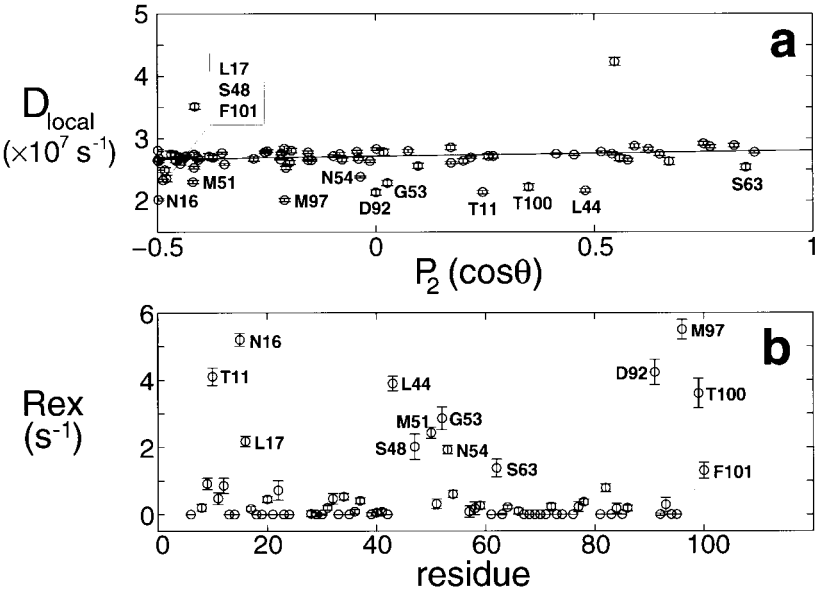

Figure 4. (a) Local diffusion constants, D_{local} ($D_{local} = 1/6\tau_{m,i}$, where $\tau_{m,i}$ is the effective rotational correlation time of residue *i*; equation (1)) calculated from ^{15}N T_1 and T_2 values plotted as a function of $P_2(\cos\theta)$ [$(3\cos^2\theta - 1)/2$] where θ is the angle between the NH bond vector and the principal axis of the fitted axially symmetric diffusion tensor. The line corresponds to the predicted relationship between D_{local} and $P_2(\cos\theta)$ (Lee *et al.*, 1997): $D_{local} = D_{iso} - P_2(\cos\theta)(D_{par} - D_{perp})/3$. Labeled peaks were found to have large R_{ex} contributions t T_2 and were not included in the diffusion tensor analysis. (b) Values for R_{ex} obtained from fits of S^2, τ_e and R_{ex} to T_1, T_2 and NOE data assuming an axially symmetric diffusion tensor with $D_{iso} = 2.72 \times 10^7$ s^{-1} and $D_{par}/D_{perp} = 0.90$. Residues with $R_{ex} > 1$ s^{-1} are labeled in both (a) and (b).

model-free formalism, but it was impossible to fit the data. Dynamic light scattering was employed to establish whether the bimolecular RNA-protein complex was monomeric (i.e. one protein and one RNA), dimeric (two proteins and two RNAs) or simply contained species aggregated non-specifically and characterized by a distribution of molecular masses. Light scattering data for the free protein were consistent with a monodisperse, monomolecular sample of the molecular mass expected from the nominal mass of the protein (11 kDa). In contrast, light scattering data recorded on the bimolecular complex at 0.15-0.3 mM concentration were not consistent with a monodisperse (monomeric or dimeric) sample. The light scattering pattern is consistent instead with a sample with an effective molecular mass of 25-30 kDa, larger than expected from the nominal molecular mass (22 kDa), yet smaller than for a multimolecular aggregate. Furthermore, the light scattering data also suggested that the sample was polydisperse, i.e. it contained a distribution of molecular species. The most likely explanation for these results is that the samples of U1A in complex with RNA are polydisperse as a result of non-specific aggregation. Under these conditions, it is difficult to interpret relaxation data quantitatively (Schurr *et al.*, 1994). Therefore, we decided to analyze the relaxation data on the protein-RNA complex without using the model free formalism or other motional models. Data on side-chain dynamics of the protein-RNA bimolecular complex were also interpreted without fitting a motional model to the data.

Side-chain dynamics

Deuterium quadrupolar relaxation rates of methyl $CH_2{}^2H$ groups in both the free UIA and bimolecular complex were measured using procedures described in several publications from our group (Kay *et al.*, 1996, 1998; Muhandiram *et al.*, 1995). Representative magnetization decay curves for free U1A and for the bimolecular complex are shown in Figure 5, together with S^2_{axis} values, describing the amplitude of motion of the bond connecting the $CH_2{}^2H$ methyl with the adjacent carbon, for the free U1A protein. We chose to extract S^2_{axis} values assuming isotropic overall diffusion ($\tau_m = 6.1$ ns), since we have shown above that the degree of anisotropy in the tumbling of U1A is small. S^2_{axis} and τ_e values were obtained from T_1 and $T_{1\rho}$ data as described (Muhandiram *et al.*, 1995). The ^2H $T_{1\rho}$ values are reported for the free and bound protein in Figure 6. The sensitivity of spectra of the trimolecular complex was insufficient to allow reliable acquisition of side-chain ^2H relaxation data.

The ^2H T_1 values are widely scattered in both the free and bound U1A protein (data not shown). These values depend strongly on the rotation rate about the methyl 3-fold averaging axis. In contrast, ^2H $T_{1\rho}$ values are much more sensitive to side-chain order and these are reported in Figure 6 for

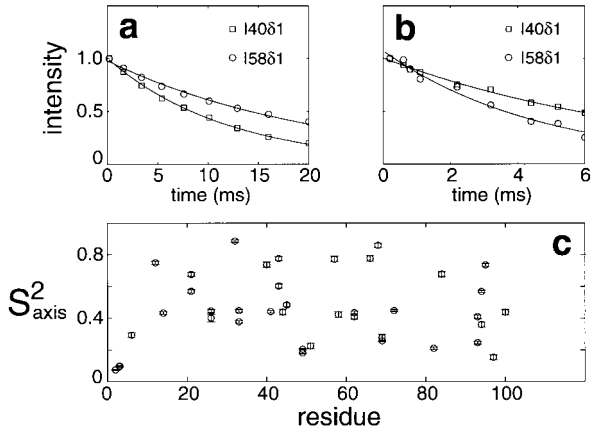

Figure 5. Normalized peak intensities as a function of relaxation delay from spectra recording the relaxation of $H_z C_z^2 H_y$ for (a) the free protein and (b) the bimolecular complex. (c) Values of the methyl axis order parameter (S^2_{axis}) for the free protein calculated assuming isotropic tumbling with a correlation time of 6.1 ns.

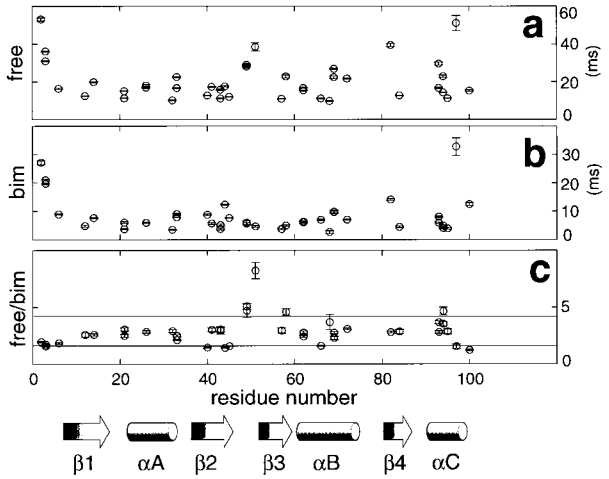

Figure 6. ^{2}H $T_{1\rho}$ values of the methyl groups of (a) free U1A and (b) U1A in the bimolecular complex. (c) The ratio of $T_{1\rho}$ values ($T_{1\rho free}/T_{1\rho bimolecular}$). The shaded region includes residues with ratios within one standard deviation of the mean. Th protein secondary structure is schematically shown at the bottom of the Figure.

the free protein and the bimolecular complex. A wide range of ^{2}H $T_{1\rho}$ values is observed for the free U1A protein, as expected from obvious differences in side-chain length and conformational constraints for individual methyl groups. In the free protein, ^{2}H $T_{1\rho}$ values range between 10-20 ms for residues buried in the hydrophobic core to 10-50 ms for solvent-exposed methyl-carrying side-chains. The S^2_{axis} values obtained for the free protein (Figure 5(c)) show the previously noted trend (Kay *et al.*, 1996; LeMaster & Kushlan, 1996; Mittermaier *et al.*, 1999) of decreasing with increasing covalent separation from the backbone. Excluding flexible termini, mean values for Ala, Thr, Ile$^{\gamma 2}$, Val, Ile$^{\delta 1}$, Leu and Met are 0.827, 0.608, 0.575, 0.526, 0.518, 0.331 and 0.260, respectively. The ^{2}H $T_{1\rho}$ values are shorter but also much more narrowly spread in the U1A-RNA complex, with most values ranging between 3-8 ms. Exceptions include residues from flexible regions at either end of the protein domain (Val3-Thr6 at the N terminus and Thr100 and Val102 at the C terminus).

Side-chain dynamics are greatly influenced by the length of the side-chain itself. In order to obtain a view of the changes observed in U1A ^{2}H $T_{1\rho}$ values upon complex formation that is least sensitive to this factor, we calculated the ratio between ^{2}H $T_{1\rho}$ values in the free and bound protein (Figure 6(c)). Side-chains displaying unusually small or large changes in ^{2}H $T_{1\rho}$ values upon complex formation (the variation being more than one standard deviation from the average calculated for all methyl groups) are explicitly shown and color-coded in Figures 7 and 8. Side-chains with methyl groups displaying unusually small changes in ^{2}H $T_{1\rho}$ values upon RNA binding are located in the flexible N-terminal end of the protein. Several residues within β2 (Ile40, Leu44 and Val45) also display anomalously small ratios of ^{2}H $T_{1\rho}$ values. Residues for which deuterium relaxation behavior changes the most are colored in red. These invariably map to the RNA binding surface and to amino acids (Ile58 and Ile94, with Ile93 and Ala95 just within one standard deviation of the mean) that pack helix C against a small hydrophobic patch that forms upon binding (Allain *et al.*, 1996). The most significant larger than average changes in ^{2}H $T_{1\rho}$ ratios are observed for two residues (Leu49 and Met51) deeply buried at the RNA-protein interface (Figure 8(a)) and located in a region of the protein (loop 3) that is crucial for specificity (Allain *et al.*, 1997). In contrast, residues with smaller than normal ratios map to a largely exposed patch on the edge of the RNA-protein interface (Figure 8(b)). The implications of these results are described in the Discussion.

Discussion

Human U1A protein binds the polyadenylation inhibition element from the 3′-untranslated region of the U1A pre-mRNA (Figure 1) with sub-nanomolar affinity and high specificity. We have previously reported the structure of U1A protein (Avis *et al.*, 1996), of one the two internal loop binding sites within PIE RNA (Gubser & Varani, 1996) and of the protein-RNA bimolecular complex (Allain *et al.*, 1996, 1997). We have also recently completed the structure of the trimolecular complex of 38 kDa containing two U1A molecules bound to the complete PIE RNA (L.V. *et al.*, unpublished results). These studies and two related crystallographic structures (Oubridge *et al.*,

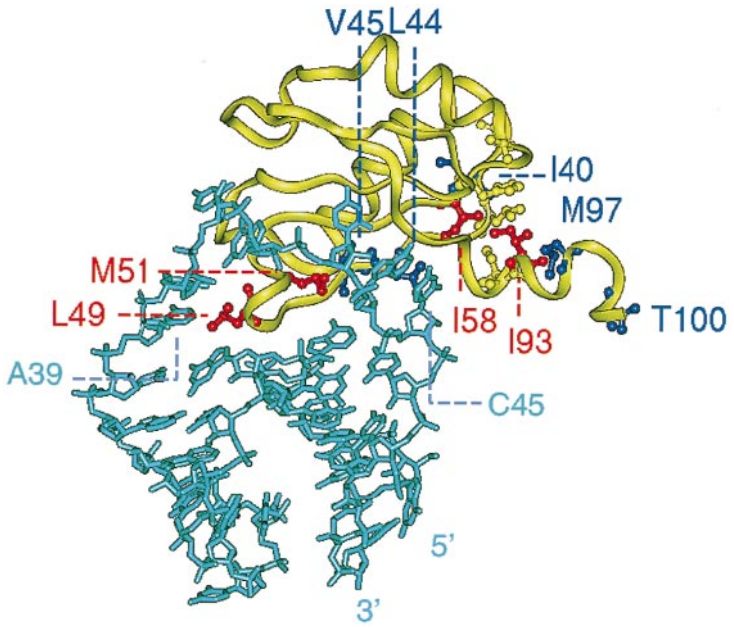

Figure 7. Changes in side-chain methyl group ^2H $T_{1\rho}$ values ($T_{1\rho free}/T_{1\rho bimolecular}$) mapped on the structure of the bimolecular complex (Allain *et al.*, 1997). Protein side-chains have been colored to reflect differences in the ratios of ^2H $T_{1\rho}$ values. Blue side-chains correspond to those with smaller than average $T_{1\rho}$ ratios, wheres red side-chains have larger than average $T_{1\rho}$ ratios. In both cases, only residues with changes larger than one standard deviation from the average (i.e. outside the shaded region in Figure 6(c)) are shown explicitly and labeled. Residues forming the hydrophobic patch positioning helix C are shown in yellow.

1994; Price *et al.*, 1998a), have provided important insight into RNA recognition by the U1A protein and other members of this large protein superfamily. However, many questions related to how U1A (and other RNPs) discriminate between different RNAs remain unclear. In order to advance our understanding of this paradigm in RNA-protein recognition, we have compared backbone and side-chain dynamics of human U1A protein in its free and RNA-bound states.

The primary data were of high quality (Figure 2) and accurate ^{15}N relaxation parameters were obtained for backbone amide resonances in the free U1A protein and in its RNA complex, as well as in the trimolecular complex with the complete PIE RNA (Figure 1). The ^2H relaxation parameters (Figure 5) were also extracted for $CH_2{}^2H$ side-chain methyl groups in the free protein and PIE-U1A complex; the molecular mass of the trimolecular complex was too large to allow extraction of reliable ^2H relaxation parameters. The relaxation data for free U1A protein could be analyzed quantitatively as described in the text, but it proved impossible to analyze the relaxation data of the protein-RNA complexes using the same approach. Light scattering data provide strong evidence that the U1A-RNA complex is polydisperse even at concentrations several fold lower than those used for NMR, very likely as a consequence of non-specific aggregation. Since aggregation significantly complicates the quantitative analysis of relaxation parameters (Schurr *et al.*, 1994), we decided to analyze relaxation data for the protein-RNA complexes focusing only on differences in the relaxation parameters (^{15}N T_2 and ^2H $T_{1\rho}$).

RNA recognition involves four regions of U1A protein: the β-sheet surface, loop 1, loop 3 and the loop connecting the end of the domain (β4) with the C-terminal helix C. Residues from the β-sheet surface, loop 3 and the loop following β4 are involved in recognition of the seven single stranded nucleotides that constitute the primary recognition site. Loop 1 and loop 3 interact also with the terminal base-pairs of the double-stranded regions of the RNA and the negatively charged phosphates in the double helical stem 2 region (Allain *et al.*, 1996, 1997). Many backbone amides from these same regions of U1A protein display anomalous relaxation behavior in the free protein, as revealed by exchange broadening of backbone NH resonances, shorter than average T_2 values and large R_{ex} contributions (Figure 4(b)). Anomalous relaxation was observed for Thr11, Asn16 and Leu17 within β1; Leu44 within β2, Ser48 and Met51 in loop3, Gly53 and Asn54 in β3; Thr89 and D90 in the loop connecting β4 with helix C and Asp92, Met97, Thr100 and Phe101 in helix C itself. Loop 3 and the region immediately following the end of the domain, from Lys88 to Lys98, play a particularly important role in determining the specificity of U1A. These are the sites of greatest genetic diversity within the superfamily (Kenan *et al.*, 1991), and mutations in these regions severely affect RNA binding. The relationship between the function of these two critical regions of U1A-RNA interaction, and the binding-induced changes in dynamics, as revealed by the present study, is discussed below.

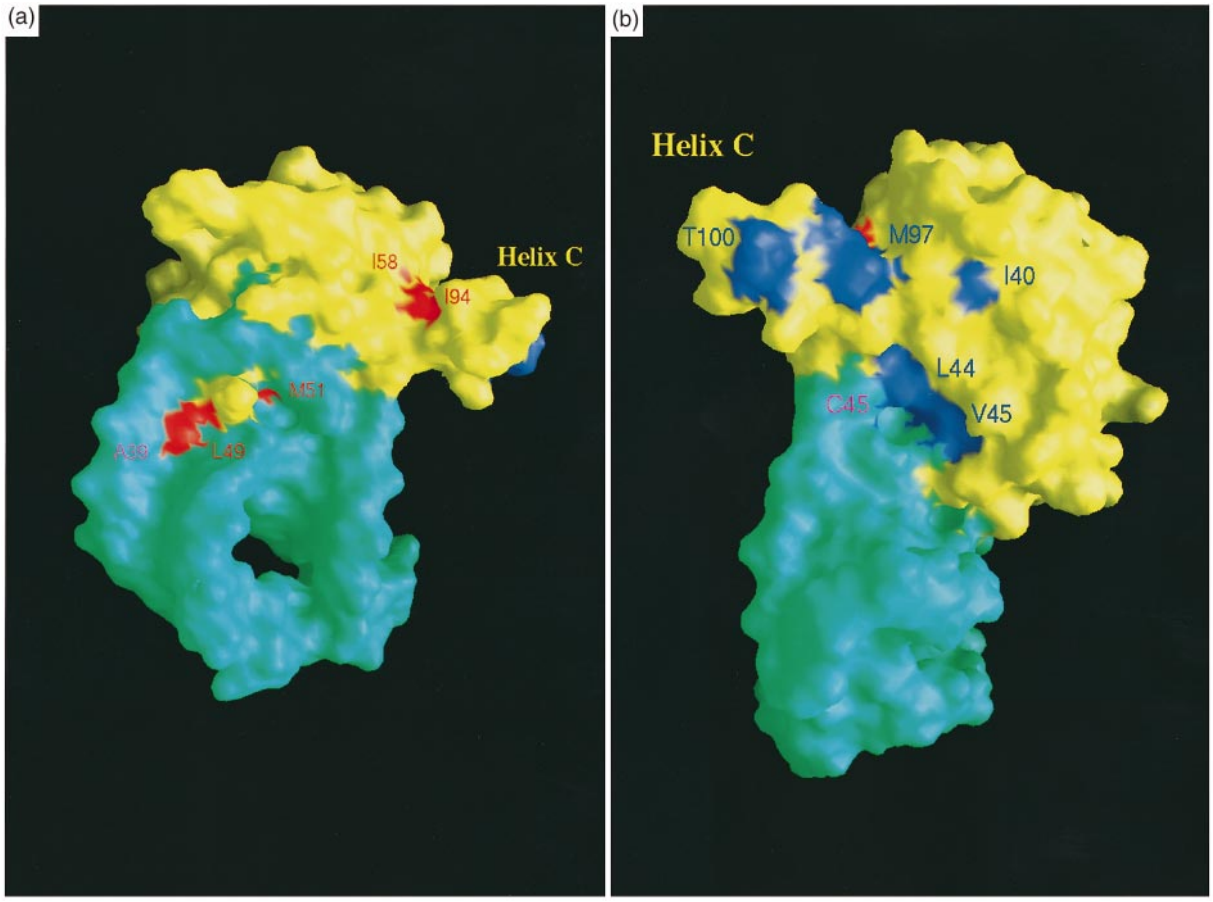

Figure 8. Surface representation of the U1A complex (Allain *et al.*, 1997). (a) Residues that become significantly more rigid in the complex (red) cluster in two distinctive patches. Leu49 and Met51 are buried at the intermolecular iterface, whereas Ile58 and Ile94 are in the hydrophobic patch that positions helix C. (b) Residues retaining significant conformational flexibility (dark blue) are found in the solvent-exposed surface of helix C (Met97 and Thr100) and in a solvent-exposed patch on the edge of th RNA-protein interface (Leu44 and Val45).

A balance between rigidity and flexibility within loop 3 provides a compromise between high specificity in nucleotide recognition and the requirement to minimize entropic losses

In the structure of the complex, loop 3 protrudes through the hole defined by the single-stranded nucleotides and the double helical stems of the RNA (Figures 7 and 8(a)). Contacts involving Ser46, Ser48, Leu49, Met51 and Arg52 from loop 3 lock the protein into this hole. The relaxation data indicate that loop 3 experiences conformational exchange in the free protein, and that these dynamic processes are quenched in the RNA complex (Figures 3 and 4(b)). The stabilization of the conformation of loop 3 results in the formation of a short, one-turn helix in the complex (Allain *et al.*, 1997). Flexibility of loop 3 in the free U1A protein could be a critical feature of U1A-RNA recognition, allowing different conformational states to be accessed during molecular recognition by induced fit. We have previously reported that five of the seven single-stranded RNA nucleotides that interact with the protein are also in conformational exchange in the free RNA (Gubser & Varani, 1996). Taken together, these results demonstrate that the U1A-PIE interaction leads to significant reduction in the conformational freedom of the regions of protein (loop 3) and RNA (the single-stranded loop) that are the primary source of specificity. Since conformational flexibility within loop 3 was also reported for another RNP protein, Musashi 1 (Nagata *et al.*, 1999), this could be a general feature of RNP-RNA recognition.

The ^{2}H $T_{1\rho}$ values of Leu49 and Met51 methyl groups within loop 3 indicate that binding leads to an unusually large decrease in side-chain flexibility as well. Residue Leu49, Met51 and Arg52 are at a center of a network of interactions that are crucial in docking loop 3 against the structure formed at the junction between the RNA helices and the internal loop (Figure 7). Mutations of the loop-closing C·G base-pairs or of A39 reduce the affinity of U1A for its specific substrate by >1000-fold (Hall, 1994) by disrupting the interactions involving Leu49, Met51 and Arg52. These interactions are essential to define the relative position of the

double helical stems (recognized through electrostatic interactions between basic residues in loops 1 and 3 and the phosphodiester backbone of stem 2 (Allain *et al.*, 1997)) and the single-stranded nucleotides (recognized by extensive interactions involving the protein β-sheet surface, loop 3 and helix C). The increase in rigidity for these side-chains is very likely to reflect the formation of a tightly packed interface (Figure 8(a)) where multiple intermolecular interactions ensure high specificity.

In contrast to what was observed for loop 3, the decrease in ^{2}H $T_{1\rho}$ values is much smaller than average for residues close to C45 (Leu44 and Val45), suggesting that these side-chains retain significant conformational flexibility in the complex. In the structure, these residues are at the edge of the protein-RNA interface and partially solvent exposed (Figure 8(b)). While substitutions of individual nucleotides near the 5′-end of the RNA single-stranded loop (A39 and G42, for example) lead to large reductions in affinity, C45 at the 3′-end of the loop can be mutated, even to a guanine base, without a large loss of binding energy. The relative insensitivity to mutations of interfacial nucleotides near the 3′-end of the single-stranded loop had been proposed to arise from residual local flexibility (Allain *et al.*, 1997). This suggestion is strongly supported by the results of deuterium relaxation experiments reported in Figure 6 and graphically presented in Figures 7 and 8(b). Relaxation studies of protein-protein interfaces have suggested that a compromise is necessary between maximal specificity (at large entropic cost) and complete lack of selectivity (Kay *et al.*, 1996). Conformational flexibility of Leu44 and Val45 may reduce the ability to discriminate between different nucleotides in this region of PIE RNA, but at the same time it reduces the entropic cost associated with the loss of conformational freedom.

Conformational rearrangement within the region C-terminal to the domain is important for RNA discrimination

Sequences flanking the RNP domain are required for RNA binding in many RNP-containing proteins. In particular, the region immediately following the end of the RNP domain of U1A is essential for RNA binding. A protein construct truncated at residue 91 does not bind RNA at all (Scherly *et al.*, 1991), truncation at residue 95 reduces binding 30-fold (Hall, 1994; Jessen *et al.*, 1991; Scherly *et al.*, 1991) and substitution of Lys96 and Lys98 with Gln reduce affinity significantly (Jessen *et al.*, 1991). Residues 91-98 form a well-defined α-helix (Allain *et al.*, 1996; Avis *et al.*, 1996) and residues 88-92 are involved in extensive intermolecular interactions with nucleotides C43, A44 and C45 (Allain *et al.*, 1996; Oubridge *et al.*, 1994). Helix C lies in different positions in free and RNA-bound U1A protein (Figure 9, top). In the free protein, helix C lies across the surface of the β-sheet and covers a large part of the RNA-binding surface (Avis *et al.*, 1996). In the complex, its position is defined by hydrophobic interactions between Ile93, Ile94 and Met97 from helix C with His10, Leu41, Ile58 and Val62 (Figure 7).

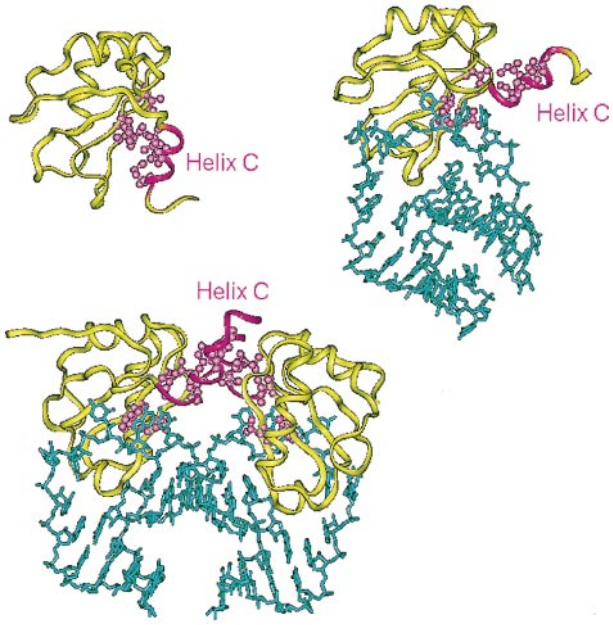

Figure 9. Structure of free U1A protein (upper left) (Avis *et al.*, 1996) and of the bimolecular complex between U1A and the half-site RNA regulatory element (upper right) (Allain *et al.*, 1996, 1997). The structure of the trimolecular complex between two U1A molecules and the complete PIE RNA is shown at the bottom (L.V. *et al.*, unpublished results). The side-chain of hydrophobic residues that position helix C in each complex are explicitly shown and are colored purple.

We proposed that helix C may contribute to molecular discrimination through two mechanisms (Allain *et al.*, 1997): First, through the establishment of interactions involving the β4-helix C loop. Second, by shielding the RNA-binding surface of U1A protein from non-specific RNA molecules. Interactions between U1A and non-cognate RNAs may provide insufficient energy to drive the conformational change in helix C, thereby reducing the affinity for non-cognate RNAs. The present data lend support to the important role of helix C realignment in RNA binding and discrimination. Residues within helix C and at the junction between helix C and the end of the RNP domain are characterized by significant conformational exchange contributions to ^{15}N relaxation (Figure 4(b)). This junction is the site of a sharp conformational rearrangement during the repositioning of helix C away from the β-sheet surface upon RNA binding (Figure 9, top) (Allain *et al.*, 1996). Residues that contribute to the formation of the hydrophobic patch that positions helix C in the complex, also show sig-

nificantly larger than average decreases in ^2H $T_{1\rho}$ values upon complex formation (Figure 7). In the construct studied here (2-102), helix C moves as a rigid unit between conformationally distinct states (Avis *et al.*, 1996), and this is the likely origin of the chemical exchange observed in this study. In contrast, the orientation of helix C in a longer fragment of U1A studied by us in the past (2-117) is much better defined. The longer construct does not show the same pattern of line broadening in the region linking the core U1A domain with helix C, suggesting that the micro- to millisecond time-scale motions present in this region in the shorter construct are absent in U1A (2-117) (Avis *et al.*, 1996). This observation is significant, because U1A (2-117) discriminates between cognate and non-cognate RNAs more effectively than the shorter construct studied here (U1A2-102) (Scherly *et al.*, 1991). A thorough analysis of the role of the residues 103-117 in improving discrimination would require a structure of the complex involving the U1A (2-117) construct, which is not available. However, the present data and the comparison of the structure of the two free protein constructs strongly suggest a mechanism by which this region of the domain could contribute to U1A specificity. Fraying of helix C and motion at the junction between helix C and the β-sheet may allow non-cognate RNAs to displace the helix more easily and bind to U1A more tightly. Displacing a more rigid helix C, such as that found in the U1A (2-117) construct, would be more costly energetically, thereby allowing better discrimination against non-cognate RNAs.

Protein-protein interactions extend helix C and increase its rigidity

RNA-protein recognition is the first step during the autoregulation of U1A expression, which relies on an RNA-dependent interaction between U1A protein and the enzyme responsible for the formation of the mature 3′-end of most eukaryotic mRNAs, poly(A) polymerase. The complete PIE RNA regulatory element contains two U1A binding sites separated by four base-pairs (Figure 1). Interaction with poly(A) polymerase and regulation of polyadenylation require cooperative binding of two proteins (Boelens *et al.*, 1993; Gunderson *et al.*, 1994, 1997; van Gelder *et al.*, 1993), and the recently completed structure of the 38 kDa trimolecular RNA-protein complex has revealed the basis of cooperativity (L.V. *et al.*, unpublished results). As illustrated in Figure 9 (bottom), U1A-U1A interactions in the cooperative complex are mediated by hydrophobic side-chains within helix C from each U1A monomer. The ^{15}N relaxation data on the trimolecular complex show that formation of this protein-protein interface selectively increases the rigidity of helix C relative to the rest of the domain (Figure 3(c)). Since cooperativity is necessary for regulation (Gunderson *et al.*, 1997; L.V. *et al.*, unpublished results), protein-protein interactions and increased rigidity in helix C, as observed here, are central the ability of U1A to regulate its own expression.

Conclusions

We have investigated the dynamics of backbone and side-chain atoms of human U1A protein free and in complex with RNA. The results presented here provide new insight into the mechanism by which specificity is ensured during RNA recognition by U1A and, presumably, other RNP. Binding induces a significant reduction in side-chain flexibility and quenching of conformational exchange in regions of the protein that are critical for specificity. The entropic penalty associated with the reduction in conformational freedom for these residues is compensated, to some extent, through the preservation of flexibility in other regions of the interface which are less critical in providing specific intermolecular contacts. This residual flexibility relaxes specificity (individual nucleotides can be exchanged without large changes in binding energy), but reduces the entropic cost associated with rigidification of protein side-chains. Thus, binding energy and discrimination in U1A-RNA recognition are modulated by balancing high rigidity and specificity in critical regions of the interface with residual mobility and relaxed specificity away from the location of the most critical intermolecular contacts.

Methods

Sample preparation

RNA samples for bimolecular and trimolecular complexes were prepared by *in vitro* transcription using phage T7 RNA polymerase and synthetic DNA templates. Sample preparation was as described in other publications from our group (Gubser & Varani, 1996; Price *et al.*, 1998b). *Escherichia coli* BL21(DE3) cells freshly transformed with the U1A protein expression vector were used in the preparation of deuterated protein samples. Kanamycin-resistance expression vectors (Gerchman *et al.*, 1994) provided more consistent expression levels, when compared with ampicillin-resistance vectors, through tighter regulation of protein expression and increased plasmid stability. A single colony from these plates was streaked on deuterated LA medium plates (DLA) and grown for about 15 hours. Deuterated media were not autoclaved to minimize the loss of ^2H_2O by evaporation. A single colony from these plates was then inoculated into 5 ml of deuterated M9 media (^{2}HM9), with the desired percentage of ^2H_2O (50 %) and $^{15}NH_4Cl$ and [^{13}C] glucose as sole sources of nitrogen and carbon, respectively. This starting culture was grown at 37 °C for about 15 hours, then added to 500 ml of pre-warmed ^{2}HM9 in a two liter flask and grown at 37 °C with vigorous shaking. The cells were induced with IPTG when the optical density reached 0.6-0.8 at 600 nm and were harvested 3 1/2 hours after induction. Cells grew approximately two to three times slower in ^{2}HM9

compared to normal M9 media, but expression levels were comparable to those obtained with non-deuterated media. Protein yields were reduced only at levels of deuteration higher than 70%. Protein purification was as previously reported (Howe *et al.*, 1998). Samples consisted of approximately 1 mM protein and RNA dissolved in 90% H_2O/10% ^2H_2O containing 10 mM phosphate buffer at pH 6.

Data collection and analysis

The ^{15}N T_1, T_2 and NOE values were recorded at 27 °C using previously published pulse schemes (Farrow *et al.*, 1994) on a Varian Unity 600 MHz spectrometer equipped with a triple-resonance, pulsed-field gradient probe with an actively shielded z gradient and z gradient amplifier. Steady-state NOE values were obtained from 2D ^1H-^{15}N correlation spectra with three seconds of ^1H saturation and a four second delay between scans and without ^1H saturation using a seven second delay between scans. T_1 values were measured from spectra recorded with values of the relaxation delay of 11, 67, 133, 211, 300, 400, 522, 666 ms, 11, 96, 210, 321, 453, 608, 774, 973, 1205 ms and 11, 133, 265, 984, 1240, 1460 ms for the free protein, bimolecular and trimolecular complexes, respectively. T_2 values were measured from spectra recorded with values of the relaxation delay of 16, 33, 49, 66, 82, 99, 115, 148 ms, 8.2, 16, 25, 33, 41, 49, 58, 66 ms and 8.2, 16, 25, 33, 41, 49, 58 ms for the free protein and bimolecular and trimolecular complexes, respectively. All spectra were recorded as complex (t_1, t_2) matrices of 128,576 points with spectral widths of 9000.9 Hz in the proton dimension and 1500, 1821, and 1882 Hz in the ^{15}N dimension for the free protein, bimolecular and trimolecular complexes, respectively. Data were processed using nmrPipe software (Delaglio *et al.*, 1995) with Gaussian apodization in both dimensions. For T_1 and T_2 experiments, peak volumes were obtained using nmrPipe software (nlinLS routine) and used to fit a two parameter function of the form $I(t) = I_o e^{-t/T1,2}$ by least-squares nonlinear regression using a conjugate gradient algorithm (Vetterling *et al.*, 1988). Errors in relaxation rates were estimated by Monte Carlo analysis (Kamith & Shriver, 1989). Steady-state NOE values were determined from the ratios of peak heights with and without proton saturation. Errors in peak height were estimated from the root-mean-square value of background noise regions.

Local values of τ_m were calculated for backbone NH bond vectors of the free protein from T_1 and T_2 values assuming fast, small amplitude internal motion (Kay *et al.*, 1989) using the r2r1_tm program available from the Palmer group web site. Using the quadric_diffusion program (Bruschweiler *et al.*, 1995), also from the Palmer group web site, an anisotropic diffusion tensor was fit to the τ_m values. Seven residues possibly experiencing chemical exchange were identified by calculating:

$$A = \left| \frac{T_1 - \langle T_1 \rangle}{\langle T_1 \rangle} + \frac{T_2 - \langle T_2 \rangle}{\langle T_2 \rangle} \right|$$

and excluding residues where this value is greater than 1.5 times the standard deviation of A obtained for all residues. Residues for which NOE values were less than 0.6 were also eliminated from the analysis. The axially symmetric model was found to give a statistically significant reduction in the χ^2 value while the additional improvement afforded by a completely assymmetric model was not statistically significant. Values of S^2, τ_e and R_{ex} were calculated as described in Results. On the basis of these fits, six additional residues were identified as having chemical exchange contributions to T_2 greater than 1 s^{-1}. Values of D_{par} and D_{perp} were refit, excluding the additional residues and $D_{par}/D_{perp} = 0.90$ was obtained (*versus* $D_{par}/D_{perp} = 0.88$ from the first round of fitting). This second round of analysis identified no additional residues with chemical exchange.

^2H data collection and analysis

Relaxation of the spin-operator terms $H_zC_zD_z$, $H_zC_z{}^2H_y$ and H_zC_z (where H_z, C_z and ^2H_z are z-components of CH_2^2H methyl ^1H, ^{13}C and ^2H magnetization, respectively) were recorded using previously described pulse schemes (Muhandiram *et al.*, 1995). $H_zC_z{}^2H_z$ rates were obtained from 2D ^1H-^{13}C correlation spectra with relaxation delays of 0.05, 4.3, 8.9, 14, 20, 26, 33, 42, 52 ms and 0.05, 3.2, 7.0, 11, 17, 24 ms for the free protein and bimolecular complex, respectively. $H_zC_z{}^2H_y$ relaxation rates were obtained from spectra with values of the relaxation delay of 0.2, 1.6, 3.4, 5.4, 7.6, 10, 13, 16, 20 ms and 0.2, 0.6, 0.8, 1.1, 2.2, 3.2, 4.4, 5.2, 6.6 ms for the free protein and bimolecular complex, respectively. H_zC_z relaxation rates were subtracted from $H_zC_z{}^2H_z$ and $H_zC_z{}^2H_y$ rates to yield ^2H $1/T_1$ and $1/T_{1\rho}$ values. Data were recorded as complex matrices of 128,576 points in (t_1, t_2) and spectral widths of 5000 and 9000.9 Hz in the ^{13}C and ^1H dimensions, respectively. Processing and analysis were performed as described for ^{15}N relaxation experiments. Values of S^2 and τ_e were obtained from fits to ^2H T_1 and $T_{1\rho}$ values for the free protein assuming isotropic tumbling with a correlation time of 6.1 ns obtained from ^{15}N T_1 and T_2 measurements. Interpretation of the ^2H data in terms of motional parameters was performed employing a value for the quadrupolar coupling constant of 168 kHz and assuming tetrahedral geometry for the methyl groups (Kay *et al.*, 1996; Muhandiram *et al.*, 1995).

Acknowledgments

It is a pleasure to thank Dr Jennifer Ashurst, Dr Mark Kelly (IMP, Berlin) and Dr Kevin Gardner (University of Toronto) for suggestions in the preparation of deuterated protein samples. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (L.E.K.) and by the MRC (G.V.). L.V. acknowledges the support of a European Union training fellowship and of the MRC. L.E.K. is an International Scholar of the Howard Hughes Medical Research Institute.

References

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

Allain, F. H.-T., Howe, P. W. A., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA binding specificity of human U1A protein. *EMBO J.* **16**, 5764-5774.

Avis, J., Allain, F. H.-T., Howe, P. W. A., Varani, G., Neuhaus, D. & 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.

Boelens, W. C., Jansen, E. J. R., van Venrooij, W. J., Stripecke, R., Mattaj, I. W. & Gunderson, S. I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell,* **72**, 881-892.

Bruschweiler, R., Liao, X. & Wright, P. E. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science,* **268**, 886-889.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). nmrPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* **6**, 277-293.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and a phosphopeptide complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry,* **33**, 5984-6003.

Gerchman, S. E., Graziano, V. & Ramakrishnan, V. (1994). Expression of chicken linker histones in *Escherichia coli*-sources of problems and methods for overcoming some of the difficulties. *Protein Exp. Purif.* **5**, 242-251.

Gubser, C. C. & 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.

Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C. & Mattaj, I. W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell,* **76**, 531-541.

Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M. & Mattaj, I. W. (1997). Involvement of the carboxy terminus of vertebrate poly A polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev.* **11**, 761-773.

Hall, K. B. (1994). Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry,* **33**, 10076-10088.

Howe, P. W. A., Allain, F. H.-T., Varani, G. & Neuhaus, D. (1998). Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. Biomol. NMR,* **11**, 59-84.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991). Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447-3456.

Kamith, U. & Shriver, J. W. (1989). Characterization of the thermotropic state changes in myosin subfragment-1 and heavy meromyosin by UV difference spectroscopy. *J. Biol. Chem.* **264**, 5586-5592.

Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry,* **28**, 8972-8979.

Kay, L. E., Muhandiram, D. R., Farrow, N. A., Aubin, Y. & Forman-Kay, J. D. (1996). Correlation between dynamics and high affinity binding in an Sh2 domain interaction. *Biochemistry,* **35**, 361-368.

Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E. & Forman-Kay, J. D. (1998). Correlation between binding and dynamics at Sh2 domain interfaces. *Nature Struct. Biol.* **5**, 156-163.

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

Lee, L. K., Rance, M., Chazin, W. J. & Palmer, A. G. I. (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and $^{13}C^{\alpha}$ nuclear spin relaxation. *J. Biomol. NMR,* **9**, 287-298.

LeMaster, D. M. & Kushlan, D. M. (1996). Dynamical mapping of *E. coli* thioredoxin via ^{13}C NMR relaxation analysis. *J. Am. Chem. Soc.* **118**, 9255-9264.

Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.

Lu, J. & Hall, K. B. (1997). Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry,* **36**, 10393-10405.

Mittermaier, A., Kay, L. E. & Forman-Kay, J. D. (1999). Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *J. Biomol. NMR,* **13**, 181-185.

Muhandiram, D. R., Yamazaki, T., Sykes, B. D. & Kay, L. E. (1995). Measurement of ^{2}H T_1 and $T_{1\rho}$ relaxation times in uniformly ^{13}C-labeled and fractionally ^{2}H-labeled proteins in solutions. *J. Am. Chem. Soc.* **117**, 11536-11544.

Nagata, T., Kanno, R., Kurihara, Y., Uesugi, S., Imai, T., Sakakibara, S.-I., Okano, H. & Katahira, M. (1999). Structure, backbone dynamics and interactions with RNA of the C-terminal binding domain of a mouse neural RNA-binding protein, Musashil. *J. Mol. Biol.* **287**, 315-330.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & 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.

Price, S. R., Evans, P. R. & Nagai, K. (1988a). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature,* **394**, 645-650.

Price, S. R., Oubridge, C., Varani, G. & Nagai, K. (1998b). Preparation of RNA-protein complexes for X-ray crystallography and NMR. In *RNA-Protein Interaction: Practical Approach* (Smith, C., ed.), Oxford University Press, Oxford.

Scherly, D., Kambach, C., Boelens, W., van Venrooij, W. J. & 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.

Schurr, J. M., Babcock, H. P. & Fujimoto, B. S. (1994). A test of the model free formulas. effects of anisotropic rotational diffusion and dimerization. *J. Mag. Reson. ser. B,* **105**, 211-224.

Teunissen, S. W. M., van Gelder, C. W. G. & 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.

The *C. elegans* Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science,* **282**, 2012-2018.

van Gelder, C. W. G., Gunderson, S. I., Jansen, E. J. R., Boelens, W. C., Polycarpou-Schwartz, M., Mattaj,

I. W. & van Venrooij, E. 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. (1997). RNA-protein intermolecular recognition. *Acc. Chem. Res.* **30**, 189-195.
Varani, G. & Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing and maturation. *Annu. Rev. Biophy. Biomol. Struct.* **27**, 407-445.
Vetterling, W. T., Press, W. H., Teukolsky, S. A. & Flannery, B. R. (1988). *Numerical Recipes in C*, Cambridge University Press, Cambridge.

Edited by P. E. Wright

(Received 13 September 1999; received in revised form 15 October 1999; accepted 15 October 1999)