The NMR structure of the 38 kDa U1A protein – PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein

Luca Varani[1], Samuel I. Gunderson[2,3], Iain W. Mattaj[2], Lewis E. Kay[4], David Neuhaus[1] and Gabriele Varani[1,5]

The status of the poly(A) tail at the 3'-end of mRNAs controls the expression of numerous genes in response to developmental and extracellular signals. Poly(A) tail regulation requires cooperative binding of two human U1A proteins to an RNA regulatory region called the polyadenylation inhibition element (PIE). When bound to PIE RNA, U1A proteins also bind to the enzyme responsible for formation of the mature 3'-end of most eukaryotic mRNAs, poly(A) polymerase (PAP). The NMR structure of the 38 kDa complex formed between two U1A molecules and PIE RNA shows that binding cooperativity depends on helix C located at the end of the RNA-binding domain and just adjacent to the PAP-interacting domain of U1A. Since helix C undergoes a conformational change upon RNA binding, the structure shows that binding cooperativity and interactions with PAP occur only when U1A is bound to its cognate RNA. This mechanism ensures that the activity of PAP enzyme, which is essential to the cell, is only down regulated when U1A is bound to the U1A mRNA.

The expression of numerous eukaryotic genes is regulated by the length of the poly(A) tail at the 3' end of the mRNA. Changes in the length of the poly(A) tail of individual mRNAs during development, differentiation and in response to growth, inflammatory or mitogenic signals control the timing and intracellular localization of gene expression by altering the stability of these mRNAs or the efficiency with which protein synthesis is initiated[1]. The regulatory elements responsible for regulation are *cis*-acting RNA structures located within the 3' untranslated regions (UTRs) of these mRNAs and the *trans*-acting protein factors that bind to them.

The best understood example of regulated gene expression through control of the status of the poly(A) tail is represented by human U1A protein. Formation of the mature 3' end of the U1A pre-mRNA by polyadenylation is inhibited in the nucleus of vertebrate cells by U1A protein itself[2–5]. Two U1A molecules bind cooperatively to an RNA signal within the U1A 3' UTR termed the polyadenylation inhibition element (PIE) located at a conserved distance from the cleavage and polyadenylation element of the U1A mRNA. The trimolecular complex formed by two U1A molecules and PIE RNA interacts with the enzyme responsible for formation of the poly(A) tail, poly(A) polymerase (PAP), and inhibits its activity. Regulation of nuclear polyadenylation is increasingly being recognized as an important step in the control of gene expression[6]. Furthermore, the general properties of the U1A regulatory mechanism could represent the more common cases where regulation of poly(A) tail length occurs after the mRNA is transported to the cytoplasm. Therefore, the U1A system provides a powerful paradigm to dissect the biochemical and structural principles underlying a crucial step in gene expression and its regulation. It also provides an ideal case to study how RNA-binding proteins interact with other proteins in the ribonucleoprotein assemblies that constitute functional and structural units during gene expression. Although significant advances have recently been made in understanding the molecular basis of RNA–protein recognition[7], the biochemical and structural mechanisms by which RNA-binding proteins form protein–protein interactions remain obscure.

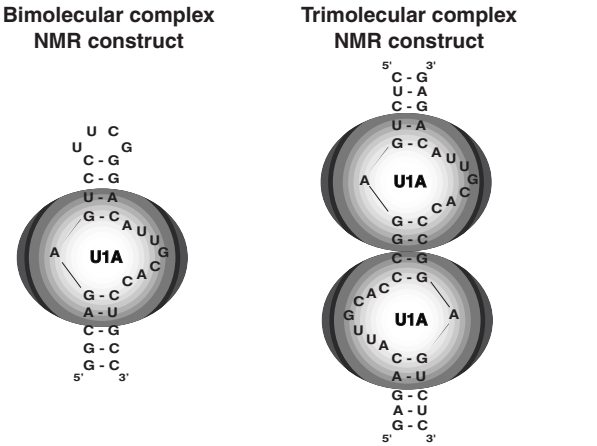

Fig. 1 RNA sequences used in NMR studies of the U1A–PIE interaction. The U1A protein is represented as a shaded sphere (see refs 8,11 for the sequence of U1A). The RNA shown here differs from the wild type sequence by the substitution of two nucleoptides within loop 1 (the lower of the two internal loops shown in the structure).

Cooperative binding is necessary for regulation of polyadenylation

PIE RNA contains a conserved secondary structure composed of

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. [2]EMBL, Gene Expression Program, 1 Meyerhofstrasse D-69117 Heidelberg, Germany. [3]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08855, USA. [4]Protein Engineering Network Centers of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

Correspondence should be addressed to G.V. *email: gv1@mrc-lmb.cam.ac.uk*

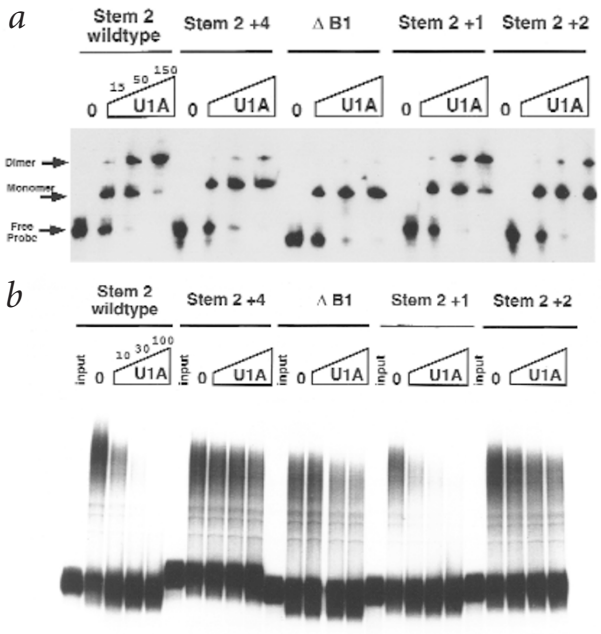

Fig. 2 Cooperativity of U1A binding depends on direct protein–protein interactions and is necessary for regulation of polyadenylation. ***a***, Electrophoretic mobility shift assays of U1A protein binding to wild type and mutant PIE RNAs. Each set of four lanes contains increasing amounts of exogenous U1A protein (as indicated in nanograms). Lanes 1–4 contain wild type PIE RNA; lanes 5–8 contain a mutant where four base pairs have been added to stem 2; lanes 9–12 contain a mutant PIE RNA known to bind only one U1A molecule (ΔB1)[3]. Lanes 13–16 and 17–20 contain mutant PIE RNAs in which stem 2 is elongated by one and two base pairs, respectively. The positions of the free RNA and U1A–RNA complexes are shown. ***b***, Effect of U1A protein on *in vitro* polyadenylation of PIE RNA. The first of each set of five lanes is the input RNA incubated with buffers in the absence of nuclear extract and U1A protein. The second lane shows polyadenylation in nuclear extract in the absence of U1A protein[2,4,5]. The next three lanes show polyadenylation in nuclear extract in the presence of increasing amounts of exogenous U1A protein (as indicated in nanograms).

two internal loops separated by a four base pair helix (Fig. 1). U1A protein monomers bind tightly to each individual internal loop, but binding of two U1A molecules to the entire regulatory element is cooperative[2,3]. The structure of the complex between the RNA-binding domain of human U1A protein (residues 1–102) and one of two conserved internal loops from PIE RNA was determined by NMR[8,9]. As the next step towards understanding the U1A regulatory system, we have studied the 38 kDa trimolecular complex formed by two U1A proteins (residues 2–102) and the complete PIE RNA (Fig. 1). Binding of two U1A proteins to PIE RNA is cooperative[3]. In order to establish whether cooperativity is due to direct protein–protein interactions, we altered the spacing between loop 1 and loop 2 by progressively extending the central stem (stem 2). Extension of stem 2 beyond the four base pairs found in the wild type sequence progressively reduced cooperative binding of two U1A proteins to PIE RNA (Fig. 2*a*). In contrast, these same extended RNAs were able to bind one molecule of U1A. As a control for where the U1A–PIE RNA complex migrates, we included the previously characterized PIE RNA mutant, ΔB1, in which one of the U1A binding sites was mutated. Thus, we conclude that PIE RNAs with an extended central stem are defective in cooperative binding of two molecules of U1A. The loss of cooperative binding abolishes inhibition of polyadenylation in the presence of PAP (Fig. 2*b*).

NMR spectral analysis

The two U1A binding sites in PIE RNA are almost symmetrical, a particularly unfavorable situation for NMR. The introduction of two mutations within loop 1 (to make it identical to loop 2) and a change of an A-U base pair in the center of stem 2 to G-C, made the complex fully symmetrical (Fig. 1). As a result of these mutations, the NMR spectral complexity was reduced, but the regulation of polyadenylation was not affected[5]. Symmetrical systems present other problems for NMR, but these can often be overcome by isotope labeling strategies based, for example, on half-filtered experiments[10]. Despite this simplification, the increase in molecular weight significantly reduces the quality of the NMR spectra. Fig. 3 compares ^{1}H-^{15}N HSQC spectra of the same U1A protein construct in the bimolecular complex with the half-site RNA regulatory element containing a single protein-binding site (Fig. 3*a*) or in the trimolecular complex with the complete PIE RNA (Fig. 3*b*). NMR relaxation measurements for backbone amides showed that ^{15}N T2 values decreased from 110–120 ms (free protein) to 40–50 ms (bimolecular complex) to 25–30 ms in the trimolecular complex[11]. The loss in spectral quality was even more severe for ^{13}C-edited spectra. TROSY experiments may have provided significant improvements in line width[12], but these were not available when most of the project was conducted. Spectral assignments were therefore based on the preparation of several fully ^{15}N- and ^{13}C labeled samples at various levels of deuteration (0%, 50%, 70% and 100%). The best compromise between the inevitable loss of signal and the increase in sensitivity due to reduced relaxation rates was obtained at 70% deuteration. Nearly complete spectral assignments were eventually obtained for both protein (98% backbone, >95% side chains) and RNA (>98% base and anomeric resonances, >90% sugar resonances).

Cooperativity in protein binding is mediated by direct protein–protein interactions and not by protein-induced changes in the RNA (Fig. 2*a*). Consistent with this, the RNA–protein interface was minimally affected by the protein–protein interactions present in the trimolecular complex. Most intermolecular protein–RNA

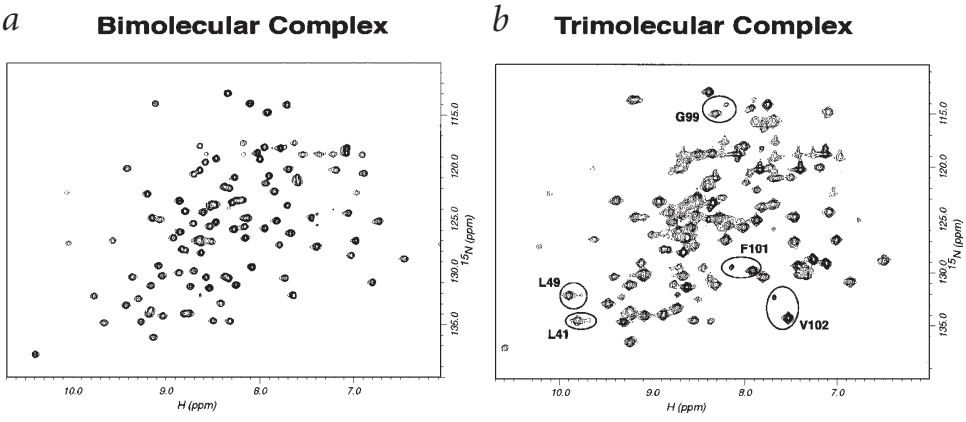

Fig. 3 NMR data for the U1A–RNA complexes. ^{1}H-^{15}N-HSQC spectra collected on samples partially deuterated to a level of 70% of ***a***, the bimolecular complex[8] and ***b***, the trimolecular U1A complex, as defined in Fig. 1.

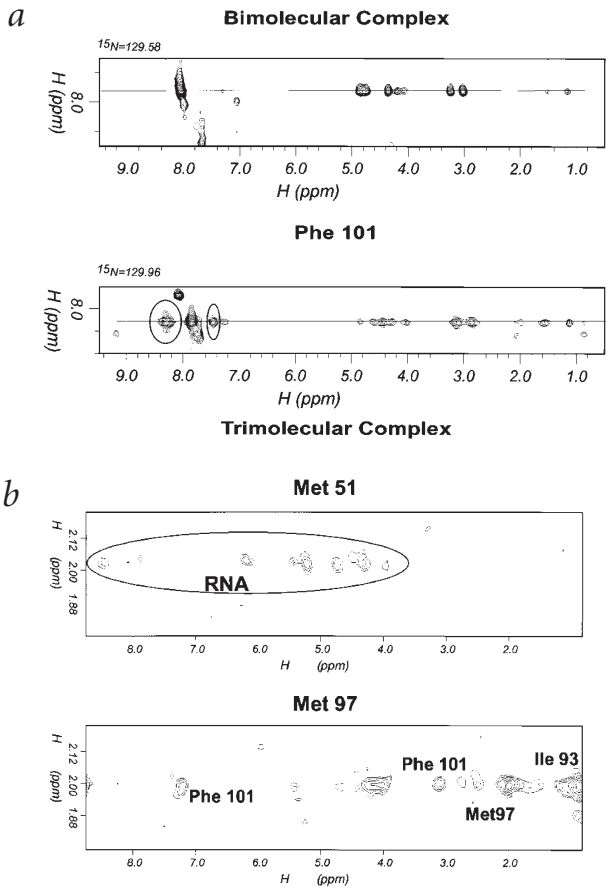

Fig. 4 NOE interactions identify helix C as the site of protein–protein interactions. ***a***, ^{1}H-^{1}H planes extracted from an ^{15}N-edited 3D NOESY experiment recorded on a sample deuterated to a level of 70% at the frequency of Phe101. Strong amide-amide NOE interactions characteristic of a helical conformation are observed in the spectrum of the homodimeric trimolecular complex (bottom), but are missing in the corresponding section of the bimolecular complex (top). ***b***, ^{1}H-^{1}H slices from a ^{13}C-1/2-filtered NOESY experiment[13] showing intermolecular protein–protein interactions involving Met 97 and protein–RNA interactions involving Met 51 methyl protons.

NOE interactions and characteristically shifted RNA and protein NMR signals were observed in the trimolecular complex. We also observed many intramolecular distance constraints involving only RNA nucleotides (Table 1). This experimental information allowed us to determine the conformation of PIE RNA and the protein–RNA interface in this trimolecular complex.

A significant proportion of the signals from the trimolecular complex studied here were similar to their counterparts in the previously studied bimolecular complex[8,9], thereby aiding assignments. However, there were also many differences (Fig. 3). The most significant changes involve amino acids in the C-terminal helix C and in regions of the protein in contact with this helix (for example His 31, Leu 41 and Ile 62). Characteristic NOE interactions between hydrophobic residues of helix C and the remainder of the RNA-binding domain demonstrated that the position of helix C remained constrained by the requirement to preserve critical interactions with the RNA. However, helix C became longer and more rigid in the trimolecular structure. The pattern of NOE interactions typical of an α-helix extends to Val 102 at the end of the construct (Fig. 4). ^{15}N relaxation measurements demonstrated that amino acids 97–100 were much more rigidly ordered in this trimolecular complex[11], with only Phe 101 and Val 102 remaining somewhat mobile.

Structure determination

The most critical information for the determination of the structure of the U1A–PIE complex were intermolecular protein–protein distances. In order to observe protein–protein contacts we recorded half-filtered experiments on a sample containing an equimolar mixture of ^{13}C-labeled and unlabeled U1A protein in complex with PIE RNA. A new experimental approach that optimizes suppression of ^{13}C-bound proton signals[13] proved invaluable, because it reduced unwanted artifacts and allowed the unambiguous observation of weak intermolecular NOE contacts. This approach led to the identification of 36 unambiguous intermolecular protein–protein distance constraints; subsequent refinement using ambiguous constraints led to the identification of a total of 66 intermolecular distance constraints. All protein–protein intermolecular NOEs involve amino acids within helix C. These results unambiguously

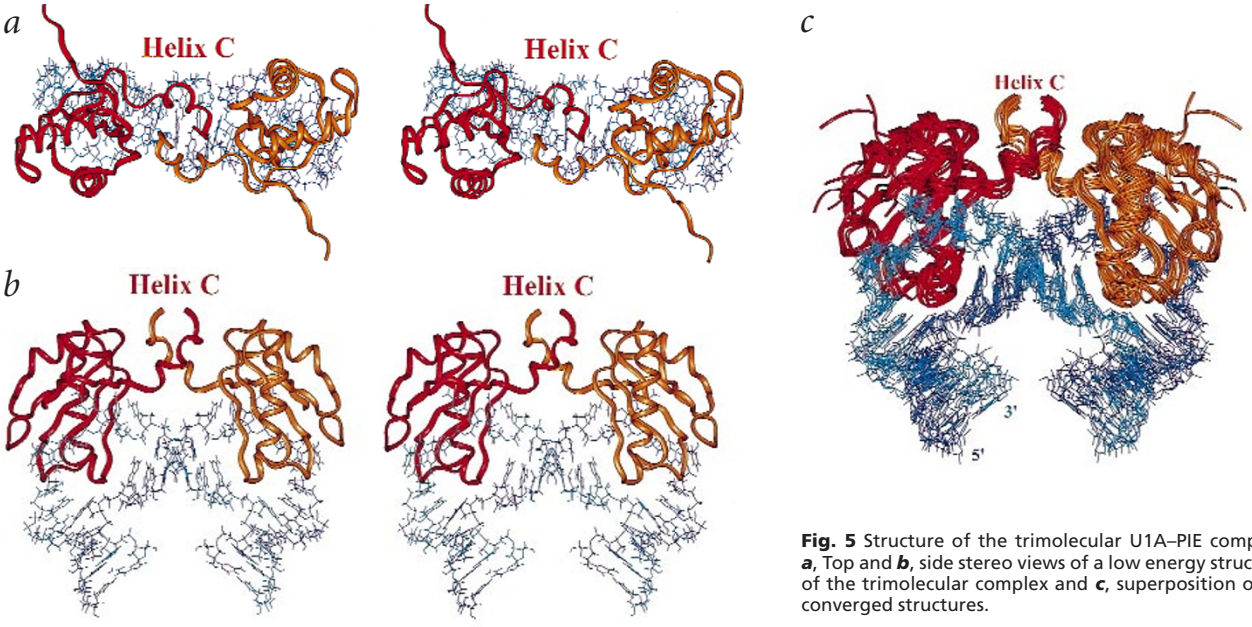

Fig. 5 Structure of the trimolecular U1A–PIE complex. ***a***, Top and ***b***, side stereo views of a low energy structure of the trimolecular complex and ***c***, superposition of 13 converged structures.

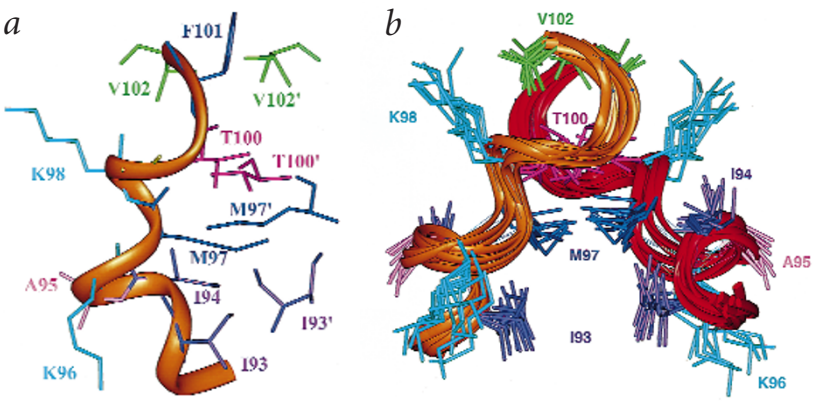

Fig. 6 Close-up view of the intermolecular hydrophobic interactions involving Ile 93, Met 97, Gly 99, Thr 100 and Val 102. ***a***, A low energy structure and ***b***, superposition of 13 converged structures.

identify the location of the protein–protein interface in the U1A structure.

We also used a new approach based on electron-proton paramagnetic relaxation to observe intermolecular long-range interactions[14,15]. We introduced single Cys mutants at three different positions in the protein, chosen to avoid disruption of the protein secondary structure and attached a nitroxide spin label to the free thiol group. The sites of substitutions were position 28 within helix α1, 61 in the loop between β3 and α2 and 103 at the C-terminal end of the construct. Substitutions at these positions resulted in proteins that were folded and well-expressed, whereas substitutions at two other sites, within the loop connecting α1 and β2, resulted in nearly complete loss of protein expression due to toxicity. The derivatized proteins induced the same unusual and characteristic chemical shifts in the RNA that were observed in the complex with the unsubstituted protein, and ^{1}H-^{15}N and ^{1}H-^{13}C HSQC spectra of the trimolecular complex were very similar to those recorded with the unsubstituted protein. Therefore, we conclude that the protein structure is not significantly affected by these substitutions. The paramagnetic label induces fast relaxation of nuclei in its vicinity. This effect is proportional to the inverse sixth power of the distance and can be observed as an increase in the line width of cross-peaks in heteronuclear correlated spectra. We prepared samples containing unlabeled PIE RNA and a 50:50 mixture of U1A protein uniformly labeled with ^{15}N and unlabeled, but containing single paramagnetic tags. Since the paramagnetically-tagged protein is not isotope labeled, heteronuclear correlated spectra only contain signals from the protein not carrying a spin-label, and any paramagnetic effect is intermolecular. We were able to obtain 30 unambiguous intermolecular distance constraints, and these were treated very conservatively by applying a single generous upper limit of 25 Å to the distance between the spin label and the paramagnetically-broadened NH signals. Because of this conservative treatment of the experimental information, the introduction of spin-labeled derived constraints had a relatively small effect on the structure, since tighter NOE constraints could be derived from half-filtered NOESY experiments. The average structures calculated with and without long-range distance constraints were superimposable within the root mean square (r.m.s.) deviation of each ensemble. However, structural statistics improved by 10–15% when long-range distances were introduced.

Structures of the trimolecular complex were calculated utilizing the same rigorous computational protocol used to define the structure of the U1A bimolecular complex[8,9,16]. Statistics of experimental constraints and of the final structures are shown in Table 1. Two views of the structure of the trimolecular U1A complex are shown in Fig. 5*a,b*, and a superposition of the converged structures is shown in Fig. 5*c*. Close-up views of the protein–protein interface are presented in Fig. 6*a* and a superposition of converged structures is shown in Fig. 6*b*. The experimental data define a structure of high precision, considering the molecular weight and the difficulties associated with studying a symmetric dimer. The r.m.s. deviation is 1.39 Å (backbone) and 1.67 Å (all heavy atoms) in both proteins, excluding the disordered N-terminal tail, as well as Phe 101 and Val 102. The overall precision for the entire complex is 2.76 Å (heavy atoms). This reflects the reduced precision of the RNA component, 2.81 Å over all atoms, due to the well-established lack of long-range distance constraints in RNA[17]. Although we have successfully used residual dipolar couplings to reduce this problem[18], we were not able to collect data of sufficient quality for the present complex. The conformation of the protein–protein interface is also well defined (Fig. 6*b*). The precision for the region of the complex comprising the helix C–helix C interface is 1.27 Å (all atoms).

The protein-protein interface is composed of helix C

The U1A–U1A interface is formed by helix C from each protein. Through this interaction, formation of the trimolecular complex presents the C-terminal end of each construct (Val 102, that immediately precedes the PAP-interacting regions, residues 102–119) on the same side of the structure, in a parallel orientation with respect to one another, and in close proximity. Helix C is amphipathic, with Asp 92, Lys 96 and Lys 98 on the RNA-facing side and several hydrophobic residues (Ile 93, Met 97,Gly 99, Thr 100 and Val 102) on the opposite surface (Fig. 6*a*). The hydrophilic face points towards the RNA in the complex and remains partially exposed to solvent. Instead the hydrophobic surface of helix C constitutes the protein–protein interface that is buried in the trimolecular complex. Formation of the U1A–U1A interface extends the hydrophobic core that holds helix C in place and buries hydrophobic residues that would otherwise be exposed to solvent. These protein–protein interactions bury ~1,100 Å^2 of surface area. By comparison, formation of the protein–RNA interface buries ~1,950 A^2 for each of the two proteins in the trimolecular complex. Therefore, the hydrophobic surface buried by the conformational change in helix C and its involvement in protein–protein interactions constitute a significant fraction of the surface area buried upon formation of the cooperative complex. Met 97 constitutes the heart of the hydrophobic patch that holds the two proteins together. It also contains Ile 93, Gly 99, Thr 100 and Val102 (Fig. 6). The two remaining hydrophobic residues of helix C, Ile 94 and Ala 95, interact intramolecularly with the rest of the protein domain (His 10, Leu 41, Leu 58 and Ile 62). Met 97 packs against Ile 93 and also interacts with Met 97, Gly 99 and Thr 100 from the neighboring protein. Packing of Met 97 against Ile 93 stabilizes the conformation of helix C. As a consequence, helix C becomes more rigid[11] and is extended by a further helical turn in the trimolecular complex (Fig. 4). Hydrogen bonds characteristic of an α-helix are observed all the way to Met 97–Thr 100, although no explicit constraint was introduced to enforce this close interaction. The side chain of Met 97 interacts with the corresponding side chain from the neighboring protein, while Gly 99, the methyl group of Thr 100 and Val 102 cap the Met 97 aliphatic side chain. In many structures, the hydroxyl

group of Thr 100 is hydrogen bonded with the backbone carbonyl of Met 97 from the neighboring protein.

How specific regulation of polyadenylation occurs

Messenger RNAs coding for spliceosomal U1A protein in vertebrates contain a regulatory region of ~50 nucleotides within their 3' UTRs, located at a conserved distance upstream of the cleavage and polyadenylation signal[2]. Two U1A molecules bind tightly and specifically to this RNA and then interact directly with poly(A) polymerase. This interaction results in inhibition of polyadenylation of the pre-mRNAs coding for U1A and down regulation of U1A expression[2–4]. U1A protein does not inhibit polyadenylation of non-cognate mRNAs even when present in very large excess. Inhibition of polyadenylation is therefore specific and depends on the presence and structural integrity of PIE RNA and on the ability of U1A protein to bind to it[2,3]. Two molecules of U1A must bind cooperatively to PIE RNA for efficient inhibition of polyadenylation to occur (Fig. 2*b*) and the cooperativity depends on direct interactions between U1A monomers (Fig. 2*a*). U1A–U1A interactions occur when the protein is bound to RNA and when the binding sites on PIE RNA are separarated by four base pairs. Increasing the length of the double helical stem separating loop 1 from loop 2 progressively reduces cooperative binding and inhibition of polyadenylation (Fig. 2). The structure of the trimolecular complex shows that increasing the separation between the two U1A binding sites would pull the two U1A proteins apart until they could no longer interact. On the other hand, shortening the spacing could also affect protein–protein interactions, either by creating steric clash or by twisting the proteins apart from one another. Finally, changing the spacer length could affect which amino acids within U1A protein are available to interact with PAP, thereby further reducing the inhibitory effect of U1A.

The structure of the trimolecular complex formed by two U1A molecules and the entire PIE RNA shows that helix C is the site of U1A–U1A interactions. The intermolecular interface forms by extension of the hydrophobic patch that stabilizes the conformation of helix C in its RNA-bound conformation (Fig. 6). Formation of the trimolecular complex brings the PAP interacting regions[4,5] from both proteins in close proximity, on the same side of the structure and in parallel orientation. In this regard, the structure is consistent with predictions based on molecular modeling that helix C is involved in the formation of the protein–protein interface and that the conformational change in helix C plays an important role in specific regulation[19]. However, the present structure is significantly different from the previously proposed model in several ways. The prediction of specific salt bridges involving Asp 24 and Lys 96, or Lys 60 and Gln 39 is not confirmed in the structure, and indeed these residues are more than 20 Å apart. Most importantly, in the previous model, PAP-interacting peptides were not brought close together by the U1A–U1A interaction. Instead, they were on opposite faces of the protein–RNA complex and were in antiparallel orientation.

The U1A system exhibits an intriguing resemblance to interactions between homo- and heterodimeric transcription factors[20]. In particular, our structure has some similarity with the yeast MATa1/MATα2–DNA complex[21]. In both the MATa1/MATα2 and U1A systems, protein–protein interactions are mediated by the carboxy-terminal tail of the nucleic acid binding domain. This region of Matα2 immediately following the nucleic acid binding domain becomes ordered upon binding its heterodimeric partner MATa1[22] and forms a short α-helix that mediates protein–protein interactions. This is precisely what we have now observed for U1A. In both the U1A and MATa1/MATα2 structures, protein–protein contacts involve hydrophobic interactions mediated by an amphipathic helix. As in the MATa1/MATα2 case, binding of two U1A molecules to PIE RNA bends the RNA, as originally shown very convincingly by fluorescence energy transfer studies of the trimolecular complex[23]. The differences between U1A and MATa1/MATα2 systems highlight the distinct functional requirements in the regulation of gene expression. Whereas U1A binds tightly to one of the sites on PIE RNA even in the absence of cooperativity, the MATa1 and MATα2 homedomains both bind DNA with modest affinity and heterodimerization is required for tight DNA binding[24]. As discussed below, U1A–U1A interactions are required to regulate the ability of U1A protein to interact with poly(A) polymerase, rather than controlling nucleic acid recognition.

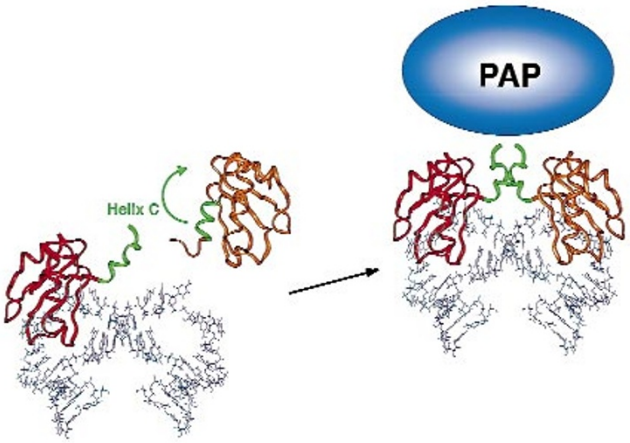

Fig. 7 The conformational change in helix C allows specific U1A–PAP interaction and selective regulation of PAP activity. Binding of U1A protein to PIE RNA changes the orientation of helix C (left) and activates the protein for homodimerization and interaction with poly(A) polymerase (right).

Cooperative binding of two U1A molecules to PIE RNA is necessary for the interaction with PAP and the inhibition of polyadenylation (Fig. 2*b*). The region of U1A involved in a direct interaction with PAP follows helix C and spans residues 103–115, although additional residues (120–137) are also highly conserved and are required for biological activity[5]. Peptides corresponding to amino acids 103–119 of U1A do not inhibit PAP in solution, but they do when conjugated to a carrier protein at sufficiently high concentration[5]. Furthermore, inhibition of polyadenylation can be observed in solution when two peptides corresponding to residues 103–115 of U1A are conjugated in parallel orientation in a single polypeptide chain, but not when these sequences are scrambled or conjugated in antiparallel orientation[25]. The sequence spanning residues 103–119 contains conserved basic and acidic amino acids spaced by approximately one α-helical turn, and mutations of some of the basic amino acids reduces inhibition of polyadenylation[25]. The present structure suggests that the U1A–U1A interface could be extended by a continuation of helix C into the 103–115 region, leaving an exposed, basic PAP-interacting surface.

Together, these results strongly suggest that the function of PIE RNA is to bring two molecules of U1A in close proximity and in the correct orientation to form productive interactions with PAP. The present structure shows that this is achieved through direct protein–protein interactions mediated by helix C. Since helix C undergoes a dramatic conformational change when U1A binds to PIE RNA[8], the PAP-interacting regions would be in very different spatial locations if the conformational change were not to take place (Fig. 7). As a consequence, the interaction between U1A and PAP

Table 1 Structural statistics[1]

Distance and dihedral angle restraints	
NOE distances	
RNA	826
Protein[2]	3,250
RNA–protein[2]	174
Protein–protein	96
Hydrogen bonds	
Protein[2]	84
RNA[3]	90
Dihedral angles	
RNA	154
Total number of experimental constraints	4,674
Structure statistics over 13 converged structures	
NOE violations	
Number > 0.2 Å	3 ± 3
Maximum violation	0.26 Å
Angle violations	
Number > 5°	0
R.m.s. deviations from mean structure (Å)	
Complete structure	
Heavy atoms	2.56 ± 0.38
All atoms	2.63 ± 0.35
Protein	
All ordered residues (Arg 7–Lys 98; Arg 7'–Lys 98')	
Backbone	1.17 ± 0.25
Heavy atoms	1.44 ± 0.24
RNA	
All residues	
Heavy atoms	2.66 ± 0.73
Single stranded loop (A13–C19; A39–C45)	
All atoms	1.37 ± 0.34
Protein–protein interface (Thr 89–Gly 99; Thr 89'–Gly 99')	
Backbone	0.81 ± 0.36
Heavy atoms	1.15 ± 0.39

[1]Root mean square (r.m.s.) deviations from ideal bond lengths, bond angles and impropers are comparable to other structures reported from our group[8,9,16].
[2]Constraint sets were duplicated for each protein in the complex and each protein–RNA interface.
[3]Including 14 weak planarity constraints for the Watson-Crick base pairs; the force constant was set at 3% of the standard X-PLOR value of 300.

can take place only once helix C has adopted its new conformation in the complex and the U1A–U1A interactions have been established as originally proposed on the basis of molecular modeling[19]. The structure described here demonstrates that U1A binding to PIE RNA ensures specificity in the regulation of U1A expression not only by the primary mechanism of direct RNA recognition, but also by promoting interactions with PAP through an RNA-induced conformational change. A distinctive feature of the auxiliary domains of RNA-binding proteins that often mediate protein–protein interactions is low sequence complexity. These domains are often composed of multiple sequences of simple two or three amino acids repeats[26,27], but protein–protein interactions mediated by these domains are often highly specific. The present structure suggests that one mechanism by which the specificity of protein–protein interactions can be modulated is by RNA-induced protein conformational changes.

Methods

Expression and purification of RNA and protein. RNA samples were prepared by *in vitro* transcription using T7 RNA polymerase and synthetic DNA templates[28]. Protein samples were overexpressed in *Escherichia coli* BL21(DE3) and purified as described[16]. Isotopically labeled samples were prepared by growing cells on M9 minimal medium enriched with appropriate labeled nutrients (^{13}C- and ^{2}D-labeled glucose and $^{15}NH_4Cl$ were from Martek; D_2O from Fluorochem). Cells were grown in highly deuterated media by selecting colonies that grew on 100% D_2O-plates overnight; no gradual adaptation was found to be necessary. Colonies were used to incubate overnight starter cultures that were then used to start large scale cultures (typically 2 l). ^{2}D concentrations quoted in this paper are nominal and refer to the D_2O/H_2O ratio in the growth medium; the actual deuteration level in the protein is 5–10% lower. Cellular growth was significantly slower in deuterium (2–3 fold), but yields of cells grown up to 70% D_2O were only slightly lower than cells grown in H_2O.

U1A mutants containing single Cys residue substitutions were purified by the same method used for the wild type protein after insertion of the desired mutation was achieved by site-directed mutagenesis (QuickChange, Stratagene) and verified by DNA sequencing. Nitroxide spin labels were attached by incubating the protein with iodoacetamido-proxyl dissolved in ethanol at room temperature. Progression of the reaction was monitored by MALDI-TOF mass spectrometry. The concentration of the protein and proxyl reactants was 0.1–0.5 mM, and complete attachment of the spin label was observed after 6 h at equimolar concentrations of protein and proxyl. Under these conditions, no reactivity with Lys side chains or protein degradation was detected by mass spectrometry (data not shown). The spin label was added at the same concentration as the protein to minimize the amount of unreacted proxyl in the NMR sample. Despite exhaustive dialysis after the reaction, it was found to be very difficult to completely purify unattached nitroxide labels. These are deleterious to the experiment, because they cause non-specific electron-proton relaxation by binding exposed hydrophobic patches in the protein.

NMR spectroscopy. RNA and protein samples were mixed after dialysis against the final buffer (10 mM Na-phosphate, 0.1 mM EDTA, pH 6) and dissolved in 300 µl in Shigemi microtubes. Most NMR experiments were recorded at 600 MHz on a Bruker DMX spectrometer equipped with triple resonance, single axis gradient, actively shielded probes. Some two-dimensional NOESY experiments were recorded at 800 Mhz on a Bruker DRX spectrometer, while the three-dimensional ^{13}C-half filtered experiments were recorded at 600 Mhz on a Varian Inova spectrometer. The sample concentration was typically 1 mM.

^{1}H-^{15}N HSQC spectra were recorded with Watergate solvent suppression[29] or with the HMQC-11 pulse sequence[30] (the latter yielded superior results for rapidly exchanging resonances). 2D NOESY and ^{1}H-^{13}C HSQC spectra were recorded in D_2O using standard pulse sequences. 2D NOESY spectra in water were recorded with either Watergate[31] or 11 solvent suppression, using mixing times of 50 and 100 ms. 3D NOESY spectra included ^{15}N-edited NOESY-HMQC-11 (mixing time 100 ms), ^{15}N-edited NOESY-HSQC (mixing time 100 ms), ^{13}C-edited NOESY-HMQC in D_2O (mixing time 50 and 100 ms) and ^{13}C-edited HSQC-NOESY (mixing time 100 ms). A 3D 1/2X-filtered NOESY experiment[13] was acquired in D_2O on a sample containing a 50%:50% mixture of unlabeled and ^{13}C-labeled U1A protein and unlabeled RNA. This experiment employed frequency-swept ^{13}C adiabatic inversion pulses to achieve optimal suppression of ^{13}C-bound proton signals in the half-filtered experiments[13]. Backbone ^{15}N T2 relaxation times are 25-30 ms[11] and TROSY experiments were not available when assignments were obtained. Therefore, we did not attempt to record triple resonance experiments. Assignments were instead obtained by analysis of 2D and, especially, 3D ^{13}C- and ^{15}N-edited NOESY spectra. This procedure made use of the similarity of many of the spin systems in this complex to the counterparts in the bimolecular complex previously assigned[9,16].

Structure determination and refinement. The statistics of the constraints used for structure calculation are reported in Table 1. The NOE-based distance constraints were supplemented with 90 hydrogen bonding and weak planarity restraints for Watson-Crick base pairs (the force constant for planarity restraints was set to 3% of the standard X-Plor value of 300, that is, 15 kcal Å^{-2}, and these were added only during the refinement/minimization stage). In addition, 84 hydrogen bonds were included within the protein for slowly exchanging backbone

amide resonances and assigned as described[16]. A total of 154 dihedral angle constraints on the RNA backbone were included for the A-form double helical regions away from the protein–RNA interfaces (stem 1 and stem 3, see Fig. 1). These constraints were based on the analysis of experimental data collected for the free RNA and the demonstration that these regions of the RNA are not affected by the presence of U1A protein. Intramolecular distance and hydrogen-bonding constraints were supplemented with 174 protein–RNA distance constraints.

A total of 66 NOE-based distance constraints between U1A proteins were derived from the analysis of half-filtered and three-dimensional NOESY experiments. We also introduced 30 long range distance constraints obtained from electron-proton relaxation experiments conducted on three protein mutants (K28C; E62C and a construct containing an additional Cys at the C-terminus of the domain). Backbone amides and side chain methyl groups close to the paramagnetic probe were identified as follows[14,15]. ^{1}H-^{15}N HSQC spectra of the complex of U1A molecules containing single Cys residues at each of the three positions (28, 61 and 103) with nitroxide spin labels were very similar to spectra of the wild type protein, when recorded under reducing conditions. This demonstrates that the mutations and derivatization with the spin label did not affect the structure of the complex. Residues in close proximity to the label were then identified by preparing samples containing a mixture of spin labeled and ^{13}C-^{15}N labeled U1A protein and by recording heteronuclear correlated spectra before and after reduction of the paramagnetic species. The unpaired electrons were quenched by reducing with an excess of ascorbate and we recorded new sets of HSQC spectra under otherwise identical conditions. Close proximity with the spin label resulting in intermolecular electron-proton interaction was identified by the observation of enhancement and sharpening of the signal of certain resonances upon reduction of the paramagnetic species. Since the spin label is attached through a relatively long and flexible linker, we did not attempt to quantify intermolecular distances, for example by measuring paramagnetic contributions to T1 and T2 relaxation times[32]. Instead, we applied a single conservative upper limit of 25 Å for the intermolecular distance between the spin label and the site of intermolecular electron-proton relaxation observed in HSQC spectra (backbone NH).

Structures for the trimolecular complex were calculated using the same rigorous approach adopted for the determination of several structures of RNA–protein complexes determined in our laboratory[8,16,33]. This procedure does not make any assumption, at any stage of the protocol, about the geometry of the RNA or protein structure, or the nature of the intermolecular interface, but simply folds the protein and RNA structures simultaneously from random starting coordinates. Non-crystallographic symmetry was imposed with a penalty set at 2% of the standard X-PLOR value, that is, 10 kcal Å^{-2}). 50 random starting structures were generated and 13 of these defined the ensemble on which the description of the structure is based. Selection of converged structures relied uniquely on the quality of the stereochemistry and the agreement with the experimental data, and was done using energy-ordered r.m.s. deviation profiles[34], as described for other structures determined by our group[16,35]. Structural statistics are reported in Table 1.

Coordinates. The coordinates for the structures have benn deposited in the Protein Data Bank (accession code 1DZ5).

Acknowledgments

We thank F. Allain and C. Gubser for their help at the initial stages of the project; M. Kelly (IMP Berlin) and K. Gardner (University of Toronto) for suggestions on the preparation of deuterated protein. L.V. acknowledges the support of an EU studentship and the MRC. We would also like to thank the Florence Large Scale facility for Biomolecular NMR, funded in part by the EU, for access to the 800 Mhz spectrometer.

Received 20 December, 1999; accepted 16 February, 2000.

1. Wickens, M., Anderson, P. & Jackson, R.J. Life and death in the cytoplasm: Messages from the 3' end. *Curr. Op. Genet. Develop.* **7**, 220–232 (1997).
2. Boelens, W.C. *et al.* The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* **72**, 881–892 (1993).
3. van Gelder, C.W.G. *et al.* 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 (1993).
4. Gunderson, S.I. *et al.* The Human U1A snRNP Protein Regulates Polyadenylation via a Direct Interaction with Poly(A) Polymerase. *Cell* **76**, 531–541 (1994).
5. Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M. & Mattaj, I.W. 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 (1997).
6. Barabino, S.M.L. & Keller, W. Last but not least: regulated poly(A) tail formation. *Cell* **99**, 9–11 (1999).
7. Varani, G. & Nagai, K. RNA Recognition by RNP proteins during RNA processing and maturation. *Ann. Rev. Biophys. Biomol. Struct.* **27**, 407–445 (1998).
8. Allain, F.-H.T. *et al.* Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**, 646–650 (1996).
9. Allain, F.H.-T., Howe, P.W.A., Neuhaus, D. & Varani, G. Structural Basis of the RNA binding specificity of human U1A protein. *EMBO J.* **16**, 5764–5774 (1997).
10. Folkers, P.J.M., Folmer, R.H.A., Konings, R.N.H. & Hilbers, C.W. Overcoming the ambiguity problem encountered in the analysis of nuclear Overhauser magnetic resonance spectra of symmetric dimer proteins. *J. Am. Chem. Soc.* **115**, 3798–3799 (1993).
11. Mittermaier, A., Varani, L., Muhandiram, D.R., Kay, L.E. & Varani, G. Changes in sidechain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.* **294**, 967–979 (1999).
12. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuation of T2 relaxation by mutual cancellation by dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
13. Zwhalen, C. *et al.* Methods for Measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/*boxb* RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721 (1997).
14. Gillespie, J.R. & Shortle, D. Characterization of long-range structure in the denatured state of Staphyloccocal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158–169 (1997).
15. Ramos, A. & Varani, G. A new method to detect long-range protein–RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).
16. Howe, P.W.A., Allain, F.H.-T., Varani, G. & Neuhaus, D. Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. Biomol. NMR* **11**, 59–84 (1998).
17. Allain, F.H.-T. & Varani, G. Structure of the P1 helix from group I self splicing introns. *J. Mol. Biol.* **250**, 333–353 (1995).
18. Bayer, P., Varani, L. & Varani, G. Refinement of the structure of protein–RNA complexes by residual dipolar coupling analysis. *J. Biomol. NMR* **14**, 149–155 (1999).
19. Jovine, L., Oubridge, C., Avis, J.M. & Nagai, K. Two structurally different RNA molecules are bound by the spliceosomal protein U1A using the same recognition strategy. *Structure* **4**, 621–631 (1996).
20. Wolberger, C. Multiprotein–DNA complexes in transcriptional regulation. *Ann. Rev. Biophys. Biomol. Struct.* **28**, 29–56 (1999).
21. Li, T., Stark, M.R., Johnson, A.D. & Wolberger, C. Crystal structure of the MATa1/MATα2 homeodomain heterodimer bound to DNA. *Science* **270**, 262–269 (1995).
22. Phillips, C.L., Stark, M.R., Johnson, A.D. & Dahlquist, F.W. Heterodimerization of the yeast homeodomain transcriptional regulator α2 and a1 induces an interfacial helix in α2. *Biochemistry* **33**, 9294–9302 (1994).
23. Grainger, R.J., Norman, D.G. & Lilley, D.M.J. Conformational consequences of binding of U1A protein to the 3' untranslated region of its pre-mRNA. *J. Mol. Biol.* **288**, 585–594 (1999).
24. Stark, M.R. & Johnson, A.D. Interaction between two homeodomain proteins is specified by a short C-terminal tail. *Nature* **371**, 429–432 (1994).
25. Klein Gunnewiek, J.M.T. *et al.* 14 Residues of the U1 snRNP-specific U1A protein is involved in homodimerization, cooperative RNA binding and inhibition of polyadenylation. *Mol. Cell Biol.* **in the press** (2000).
26. Biamonti, G. & Riva, S. New insights into the auxiliary domains of eukaryotic RNA binding proteins. *FEBS Lett.* **340**, 1–8 (1994).
27. Birney, E., Kumar, S. & Krainer, A.R. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816 (1993).
28. Price, S.R., Oubridge, C., Varani, G. & Nagai, K. Preparation of RNA–protein complexes for X-ray crystallography and NMR. In *RNA–Protein Interaction: Practical Approach* (ed. Smith, C.) (Oxford University Press, 1998), p. 48–72.
29. Mori, S., Abeygunawardana, C., O'Neil Johnson, M. & van Zijl, P.C.M. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Mag. Res.* **B 108**, 94–98 (1995).
30. Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. & Tschudin, R. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *J. Mag. Res.* **86**, 304–318 (1990).
31. Talluri, S. & Wagner, G. An optimized 3D NOESY-HSQC. *J. Mag. Res.* **B 112**, 200–205 (1996).
32. Gillespie, J.R. & Shortle, D. Characterization of long-range structure in the denatured state of Staphyloccocal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170–184 (1997).
33. Ramos, A. *et al.* RNA Recognition by a Staufen double-stranded RNA binding domain. *EMBO J.***in the press** (2000).
34. Fletcher, C.M., Jones, D.N.M., Diamond, R. & Neuhaus, D. Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR* **8**, 292–310 (1996).
35. Varani, G., Aboul-ela, F. & Allain, F.H.-T. NMR investigations of RNA structure. *Prog. NMR Spectr.* **29**, 51–127 (1996).