Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

U1A RNA-binding domain at 1.8 Å resolution

Peter B. Rupert, Hong Xiao and Adrian R. Ferré-D'Amaré*

Fred Hutchinson Cancer Research Center, USA

Correspondence e-mail: aferre@fhcrc.org

Received 7 March 2003
Accepted 21 May 2003

PDB Reference: U1A RNA-binding domain, 1nu4, r1nu4sf.

The human U1A RNA-binding domain (RBD1) adopts one of the most common protein folds, the RNA-recognition motif, and is a paradigm for understanding RNA–protein interactions. A 2.8 Å resolution structure of the unbound RBD1 has previously been determined [Nagai *et al.* (1990). *Nature (London)*, **348**, 515–520] and revealed a well defined α/β core with disordered termini. Using a longer construct, a 1.8 Å resolution structure of the unbound domain was determined that reveals an ordered C-terminal helix. The presence of this helix is consistent with a solution structure of the free domain [Avis *et al.* (1996). *J. Mol. Biol.* **257**, 398–411]; however, in the solution structure the helix occludes the RNA-binding surface. In the present structure, the helix occupies a position similar to that seen in a 1.9 Å resolution RNA–RBD1 complex structure [Oubridge *et al.* (1994). *Nature (London)*, **372**, 432–438]. The crystals in this study were grown from 2.2 *M* sodium malonate. It is possible that the high salt concentration helps to orient the C-terminal helix in the RNA-bound conformation by strengthening hydrophobic interactions between the buried face of the helix and the α/β core of the protein. Alternatively, the malonate (several molecules of which are bound in the vicinity of the RNA-binding surface) may mimic RNA.

1. Introduction

The RNA-recognition motif (RRM) is the most abundant RNA-binding domain (RBD) in vertebrates (Varani & Nagai, 1998; Hall, 2002). It is the sixth most abundant protein family in the murine genome (Mouse Genome Sequencing Consortium, 2002). This globular domain is an α/β sandwich in which two α-helices pack against an antiparallel β-sheet. The topology of the domain is $\beta\alpha\beta$-$\beta\alpha\beta$ and the RNA binds to the solvent-exposed side of the β-sheet.

The human U1A protein is part of the U1 small nuclear ribonucleoprotein particle, a large RNA–protein complex involved in pre-mRNA splicing. The U1A protein contains two RRM domains. One of these, the N-terminal RBD1, is responsible for sequence-specific RNA binding and consists of ~100 amino-acid residues.

Early NMR studies of a construct comprising residues 1–102 revealed the C-terminus to be extremely mobile (Nagai *et al.*, 1990). A 2.8 Å resolution crystal structure of unbound U1A RBD1 (Nagai *et al.*, 1990) showed a well formed compact core (residues 6–88) with two α-helices packing against a four-stranded β-sheet. Both termini were disordered, as was a loop between $\beta1$ and $\beta2$ in one of the two molecules in the asymmetric unit. Diffraction-quality crystals were only obtained after the U1A RBD1 was truncated to residues 1–95. This shortened protein exhibited reduced stability and affinity for RNA (Zeng & Hall, 1997). NMR experiments using a 117-residue U1A RBD1 construct demonstrated the existence of a C-terminal helix, helix *C* (residues 89–95; Avis *et al.*, 1996). This helix occluded the RNA-binding surface by packing across the β-sheet. However, tryptophan fluorescence measurements found helix *C* associated only weakly with the β-sheet (Jean *et al.*, 1999).

A co-crystal structure has been determined of human U1A RBD1 in complex with a 21-nucleotide RNA hairpin (Oubridge *et al.*, 1994). To obtain well ordered crystals of the complex, it was necessary to use a longer construct consisting of residues 1–98 which also included two surface mutations (Y31H, Q36R) distant from the binding surface. In this 1.9 Å resolution structure, the protein makes specific RNA contacts using the exposed surface of the β-sheet, the loop between $\beta2$ and $\beta3$ and helix *C* (residues 91–96). Solution studies showed that both helix *C* and the loop between $\beta2$ and $\beta3$ exhibit conformational changes on binding RNA (Kranz & Hall, 1998; Mittermaier *et al.*, 1999).

We report a 1.8 Å resolution structure of the unbound U1A RBD1 determined using a

Table 1
Diffraction data statistics.

Values in parentheses are for the highest resolution shell.

Resolution (Å)	30–1.8 (1.86–1.8)
Wavelength (Å)	1.54
No. of reflections	70351
No. of unique reflections	17618
Space group	*R*32
Unit-cell parameters (Å)	$a = b = 91.82$, $c = 120.45$
Completeness (%)	96.2 (76.8)
$R_{\rm merge}$† (%)	3.1 (31.1)
$I/\sigma(I)$	37.6 (2.46)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where I is an individual intensity measurement.

protein construct identical to that used in the RNA–protein complex structure. The overall fold is similar to the original unbound structure. Unexpectedly, the loop between $\beta2$ and $\beta3$ and helix *C* are ordered. Unlike the solution structure of the free domain, helix *C* occupies the same general location as in the RNA-bound structure. However, helix *C* adopts two slightly different conformations in the two RBD1 molecules in the asymmetric unit of our structure. This suggests that helix *C* can sample multiple orientations and RNA binding selects the conformation seen in the RNA-complex structure.

2. Materials and methods

2.1. Purification and crystallization

Recombinant U1A RBD1 was expressed in *Esherichia coli* and purified as described previously (Oubridge *et al.*, 1995). The protein construct encompasses residues 1–98 and contains two substitutions, Y31H and Q36R. Purified protein was dialyzed against storage buffer (10 m*M* HEPES pH 7.5, 0.1 m*M* EDTA), concentrated to $\sim$20 g l^{-1} and stored at 193 K.

U1A RBD1-(Y31H,Q36R) in conjunction with its cognate RNA hairpin-binding site has been used successfully as a module to aid in the growth of well ordered crystals of catalytic RNA molecules (Ferré-D'Amaré *et al.*, 1998; Rupert & Ferré-D'Amaré, 2001; Rupert *et al.*, 2002). However, we have found that the common precipitating agents sodium formate (Jancarik & Kim, 1991) and sodium malonate (McPherson, 2001) disrupt RNA binding by U1A RBD1. U1A RBD1 crystallizes readily under these conditions. Thus, crystals grown under these conditions contain no RNA.

We have grown a number of crystals containing only U1A RBD1 from crystallization screens of various RNA constructs that were initially complexed with the U1A protein. Typically, RNA and protein are mixed to final concentrations of 0.15 and 0.22 m*M*, respectively, in a buffer consisting of 5 m*M* potassium cacodylate pH 6.5, 1.25 m*M* $MgCl_2$ and 0.05 m*M* EDTA, and this mixture is subjected to crystallization screens. By mixing 3 µl of RNA–protein solution with 3 µl of reservoir containing 2.0–2.4 *M* sodium malonate pH 5.0, rhombohedral crystals containing only the U1A RBD1 (confirmed by denaturing polyacrylamide gel analysis of dissolved crystals) have been grown by the hanging-drop vapor-diffusion method. These protein crystals diffract X-rays to 1.8 Å resolution.

2.2. Data collection and refinement

Prior to data collection, crystals were stabilized by gradually replacing the mother liquor with a cryoprotectant solution composed of 3.1 *M* sodium malonate, 10 m*M* potassium cacodylate pH 6.5, 2.5 m*M* $MgCl_2$, 5 m*M* spermine and 10% glycerol over the course of 15 min. Stabilized crystals were mounted in loops and flash-cooled in liquid nitrogen. Diffraction data were collected on a rotating-anode X-ray generator equipped with Osmic optics and were reduced with the *HKL* package (Otwinowski & Minor, 1997). The crystals belong to space group *R*32, with unit-cell parameters $a = b = 91.82$, $c = 120.45$ Å. Data statistics are given in Table 1.

The structure was solved by molecular replacement using the original bound U1A RBD1 structure (Oubridge *et al.*, 1994) as a search model. Two U1A RBD1 molecules occupy the asymmetric unit, with a $V_{\rm M}$ (Matthews, 1968) value of 2.16 Å^3 Da^{-1}. Molecular replacement and subsequent positional and temperature-factor refinements were carried out using *CNS* (Brünger *et al.*, 1998). After correctly positioning two U1A RBD1 domains in the asymmetric unit using data from 30.0 to 4.0 Å resolution, the replacement solution gave an $R_{\rm free}$ of 43.0% and a correlation coefficient of 0.434. 10% of the reflections were set aside for cross-validation. The C-terminal helix and the loop between $\beta2$ and $\beta3$ were removed from the model and rebuilt as the electron density improved. Cycles of maximum-likelihood refinement were interspersed with manual model rebuilding using the program *O* (Jones *et al.*, 1991). After all the electron density corresponding to the protein had been accounted for, magnesium ions, water molecules and 12 well ordered malonate molecules were added to the model. In the final model, 92.3% of the residues lie in the

Table 2
Refinement statistics.

Resolution range (Å)	30.0–1.8
No. of atoms	
Total	1721
Protein	1526
Malonate	91
Ion	5
Water	99
Mean atomic displacement parameters (Å^2)	
Total	40.2
Protein	38.1
Malonate	69.1
Ion	49.9
Water	45.3
$R_{\rm work}$† (%)	20.32
$R_{\rm free}$† (%)	23.44
σ_A coordinate error (Å)	0.23
R.m.s deviations	
Bond lengths (Å)	0.0056
Bond angles (°)	1.105

† $R_{\rm work} = \sum_{hkl}\big||F_o(hkl)| - k|F_c(hkl)|\big|/\sum_{hkl}|F_o(hkl)|$. $R_{\rm free}$ is the same calculation for the randomly selected test set (10% of the data).

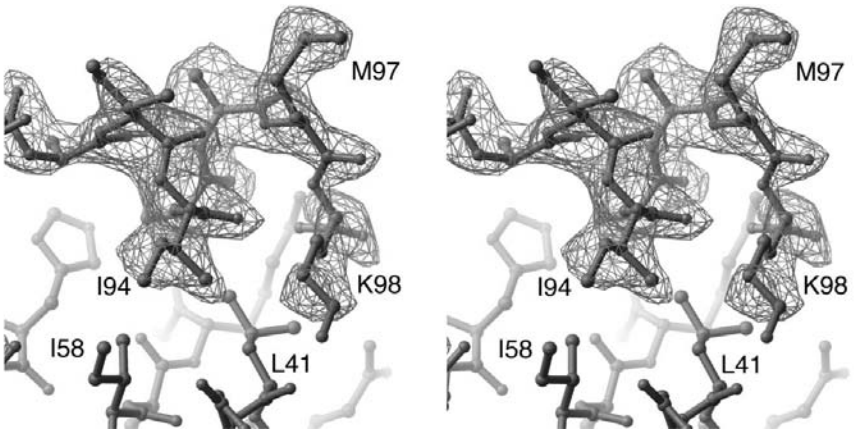
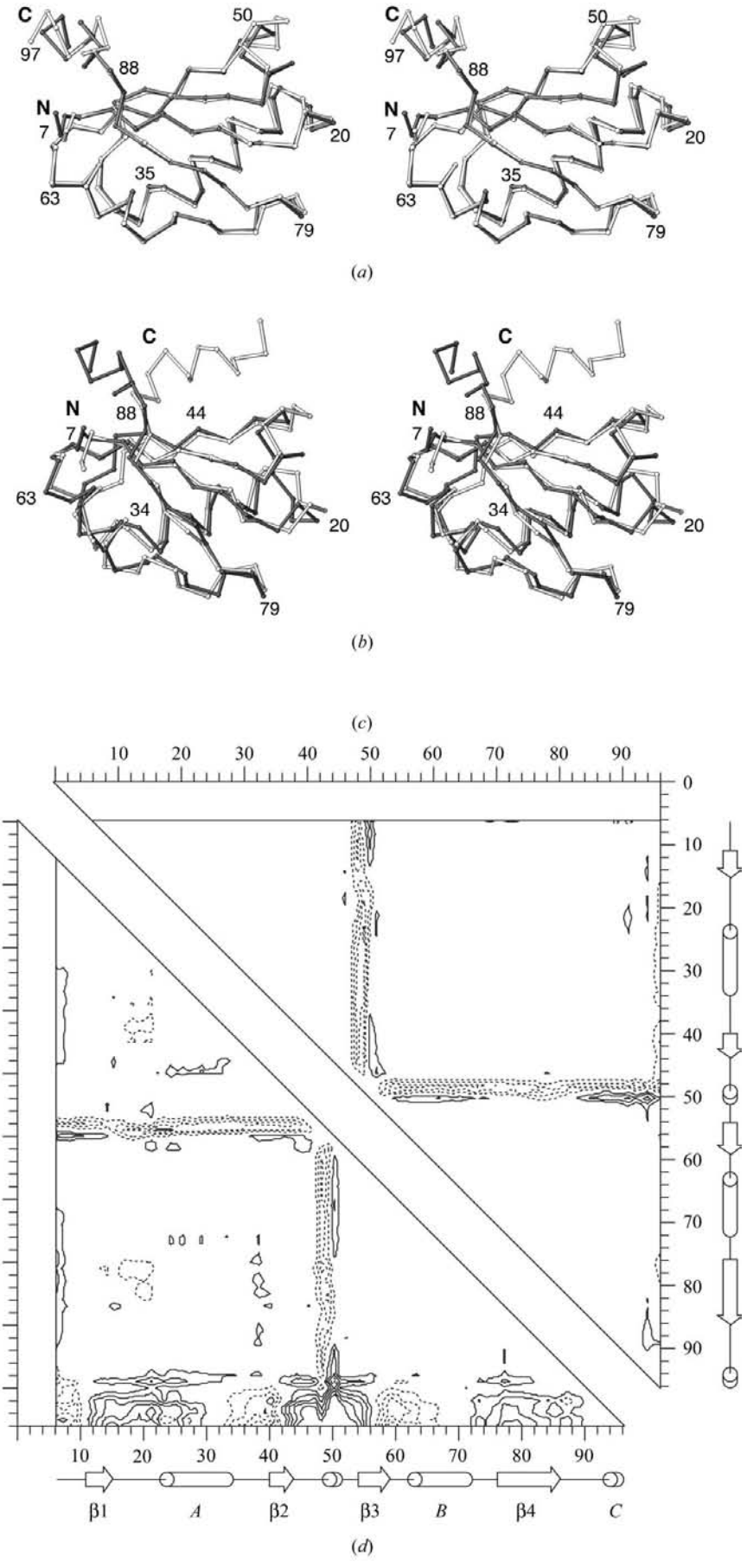


Figure 1
Stereoview of helix *C*. σ_A-weighted simulated-annealing omit electron density (contoured at 3 standard deviations above mean peak height) superimposed on helix *C* from chain *B*. Residues 89–98 were omitted from the calculation. Ile94, Ala95 and the aliphatic portion of the side chain of Lys98 pack against a hydrophobic patch on the protein core that includes residues Ile58 and Leu41. Figs. 1, 2(*a*) and 2(*b*) were prepared with the program *RIBBONS* (Carson, 1997).



most favoured region of the Ramachandran plot, with the remaining 7.7% in the additionally allowed region. Refinement statistics are summarized in Table 2.

3. Results and discussion

In our structure of the unbound RBD1, the N-terminal 5–6 residues are disordered, but the C-terminus forms an α-helix, helix *C* (Fig. 1). The structural cores of the two RBD1s in the asymmetric unit are very similar, with a root-mean-square (r.m.s.) deviation of 0.32 Å on superimposing the C^{α} atoms of residues 7–47 and 53–89. The helical register of helix *C* in the two monomers in the asymmetric unit differs by approximately 120°. In addition, the loop between β1 and β2 adopts a different conformation in the RBD1 monomers, giving an r.m.s. deviation of 3.80 Å when superimposing the C^{α} atoms of residues 45–53. The mean *B* values for helix *C* and the loop between β1 and β2 are 55.1 and 60.0 $Å^2$, respectively. These values are significantly larger than the mean *B* value of 34.4 $Å^2$ for the core. Helix *C* occupies the same general location as in the RNA-bound complex (Oubridge *et al.*, 1994) (Figs. 2*a* and 2*c*), although it has a slightly different register. The orientation of helix *C* is quite different from that in a structure of unbound RBD1 determined by NMR (Avis *et al.*, 1996) (Figs. 2*b* and 2*d*).

Our structure of RBD1 provides a novel view of how helix *C* can pack against the

Figure 2
Comparisons of U1A RBD1 structures. (*a*) Superposition of the bound structure (white) over the 1.8 Å resolution unbound structure (grey) of the same protein construct (0.46 Å r.m.s. deviation). (*b*) Superposition of the 1.8 Å resolution unbound structure (grey) over the unbound solution structure (white) of the 117-residue construct (1.11 Å r.m.s. deviation). Both superpositions were calculated with C^{α} positions of residues 8–88 (excluding helix *C*). (*c*) Difference distance matrix (Nishikawa & Ooi, 1974) plot showing the change in distance between all pairs of C^{α} atoms in the bound structure *versus* the 1.8 Å resolution unbound structure. The plot shows that the principal differences between the two structures lie in the residues between strands β2 and β3. (*d*) Difference distance matrix plot of the unbound NMR structure *versus* the 1.8 Å resolution unbound structure, showing how the major difference between the two structures is in the orientation of helix *C*. The plots are calculated by subtracting the matrix of all inter-C^{α} distances of one structure (distance matrix) from the distance matrix of the other structure. Difference distance matrices were calculated with the program *DDMP* (Richards & Kundrot, 1988). Because the matrices are symmetric, only the upper or lower triangular matrices are shown. Contours are drawn at increments of 2.0 Å, with negative contours represented by broken lines. The locations of secondary-structure elements were determined using the program *DSSP* (Kabsch & Sander, 1983).

domain. This helix is absent from the original unbound structure because the protein used to solve that structure terminated after residue 95 (Nagai *et al.*, 1990). The shorter protein fragment will not form helix *C*, but it did produce crystals that diffracted X-rays to 2.8 Å resolution. The different conformation of helix *C* observed in the unbound NMR structure (Fig. 2*b*) may be a consequence in part of the different solution conditions used. The NMR experiments were performed in a low-salt buffer containing 50 m*M* NaCl, while our crystals were grown from a buffer containing 2.2 *M* sodium malonate. Because the interface between helix *C* and the core of U1A RBD1 is hydrophobic (Fig. 1), the high-salt crystallization conditions may have stabilized this conformation. Crystal packing may also influence the position of helix *C* in our structure. Together, our crystal structure and the solution structure show that in the unbound state helix *C* is a mobile element capable of sampling multiple conformations.

In our structure of RBD1, the protein moieties that contact RNA are in the same general locations as in the complex structure (Oubridge *et al.*, 1994). Residues on the β-sheet that make direct contacts with RNA have the same conformations as in the complex. In the complex, the aromatic side chains Tyr13 and Phe56 stack against nucleotide bases of the RNA. Interestingly, in both monomers in the asymmetric unit of our structure a well ordered malonate molecule is present in the position occupied by a cytosine base in the complex. It is possible that malonate mimics RNA. Two additional elements that contact RNA, helix *C* and the loop between β2 and β3, exhibit the largest structural variability between monomers in the asymmetric unit of our structure. Helix *C* from the complex adopts a conformation more similar to that of helix *C* in one monomer (chain *A*) in the asymmetric unit of our structure, while the loop between β2 and β3 in the complex is more similar to the loop of the other monomer (chain *B*) in our structure.

Conformational rearrangements of U1A RBD1, especially of helix *C* and the loop between β2 and β3, are necessary for RNA binding (Hall, 1994; Kranz & Hall, 1998; Mittermaier *et al.*, 1999). Our unbound structure corroborates previous solution studies that showed that helix *C* is capable of forming prior to RNA binding. Our structure also shows that this helix can adopt multiple conformations.

We thank Dr Mike Quillen for assistance in producing the difference distance matrix plots. PBR was a postdoctoral trainee of the Chromosome Metabolism and Cancer training grant from the National Cancer Institute. This work was supported by the NIH (grants GM63576 and RR15943 to ARF). ARF is a Rita Allen Foundation Scholar.

References

Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K. & Neuhaus, D. (1996). *J. Mol. Biol.* **257**, 398–411.

Brünger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Carson, M. (1997). *Methods Enzymol.* **277**, 493–505.

Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. (1998). *Nature (London)*, **395**, 567–574.

Hall, K. B. (1994). *Biochemistry*, **33**, 10076–10088.

Hall, K. B. (2002). *Curr. Opin. Struct. Biol.* **12**, 283–288.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Jean, J. M., Clerte, C. & Hall, K. B. (1999). *Protein Sci.* **8**, 2110–2120.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kranz, J. K. & Hall, K. B. (1998). *J. Mol. Biol.* **275**, 465–481.

McPherson, A. (2001). *Protein Sci.* **10**, 418–422.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. E. & Varani, G. (1999). *J. Mol. Biol.* **294**, 967–979.

Mouse Genome Sequencing Consortium (2002). *Nature (London)*, **420**, 520–562.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). *Nature (London)*, **348**, 515–520.

Nishikawa, K. & Ooi, T. (1974). *J. Theor. Biol.* **43**, 351–374.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). *Nature (London)*, **372**, 432–438.

Oubridge, C., Ito, N., Teo, C. H., Fearnley, I. & Nagai, K. (1995). *J. Mol. Biol.* **249**, 409–423.

Richards, F. M. & Kundrot, C. E. (1988). *Proteins Struct. Funct. Genet.* **3**, 71–84.

Rupert, P. B. & Ferré-D'Amaré, A. R. (2001). *Nature (London)*, **410**, 780–786.

Rupert, P. B., Massey, A. P., Sigurdsson, S. T. & Ferré-D'Amaré, A. R. (2002). *Science*, **298**, 1421–1424.

Varani, G. & Nagai, K. (1998). *Annu. Rev. Biophys. Biomol. Struct.* **27**, 407–445.

Zeng, Q. & Hall, K. B. (1997). *RNA*, **3**, 303–314.