Crystallization and preliminary X-ray diffraction studies of UP1, the two-RRM domain of hnRNP A1

LANA JOKHAN,[a] AI-PING DONG,[a] AKILA MAYEDA,[b] ADRIAN R. KRAINER[b] AND RUI-MING XU[a] at [a]W. M. Keck Structural Biology Laboratory, Cold Spring Harbor, NY 11724, USA, [b]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

Abstract

The N-terminal domain of hnRNP A1 protein, termed UP1, comprises two tandem RNA-recognition motifs, both of which are necessary for efficient RNA binding and for the alternative splicing activity of hnRNP A1. Recombinant human UP1 expressed in *E. coli* has been crystallized in space group $P2_1$ with unit-cell dimensions $a = 37.94$, $b = 43.98$, $c = 55.64$ Å and $\beta = 93.9°$. The unit-cell volume is consistent with one UP1 molecule per asymmetric unit and a calculated 49% solvent content. The crystal diffraction limit is higher than 1.3 Å, and a data set to 2.0 Å has been collected. Diffraction data from one platinum and two mercury derivatives have also been collected.

1. Introduction

A1 heterogeneous nuclear ribonucleoprotein (hnRNP A1) is one of the most abundant hnRNP proteins, which associate with nascent transcripts in the nucleus (Dreyfuss, Matunis, Piñol-Roma & Burd, 1993). It has been shown to influence alternative 5′ splice site selection by antagonizing the activity of SF2/ASF and SC35, leading to the activation of distal 5′ splice sites and promoting alternative exon skipping (Fu, Mayeda, Maniatis & Krainer, 1992; Mayeda & Krainer, 1992; Mayeda, Helfman & Krainer, 1993). In addition, hnRNP A1 facilitates the efficient annealing of complementary RNA or DNA molecules *in vitro* (Kumar & Wilson, 1990; Pontius & Berg, 1990; Munroe & Dong, 1992; Portman & Dreyfuss, 1994), which may directly facilitate base-pairing interactions between snRNA's and pre-mRNA. hnRNP A1 can interact stably with U2 and U4 snRNP's *in vitro*, but not with U1 snRNP (Buvoli, Cobianchi & Riva, 1992). The interaction with U2, together with the association between hnRNP A1 and certain 3′ splice sites *in vitro* (Swanson & Dreyfuss, 1988) suggested a potential role for hnRNP A1 in the binding of U2 snRNA to the branch site in the early stages of spliceosome assembly (Buvoli *et al.*, 1992).

Human hnRNP A1 is a 34 kDa protein consisting of a single polypeptide chain of 320 amino acids. An N-terminal proteolytic fragment spanning the first 196 amino acids and known as unwinding protein 1 (UP1; Herrick & Alberts, 1976; Williams, Stone, LoPresti, Merrill & Planck, 1985) comprises two RNA-recognition motif (RRM) sequences separated by a short linker. The C-terminal region of hnRNP A1 is rich in glycine and includes several RGG repeats, which also constitute an RNA-binding motif (Kiledjian & Dreyfuss, 1992). Early studies showed that both the N-terminal UP1 domain and the C-terminal glycine-rich domain are capable of binding to nucleic acids, whereas the intact protein exhibits more

stable and highly cooperative binding, with cooperativity being the result of self-association of the C-terminal domain (Merrill, Stone, Cobianchi, Wilson & Williams, 1988; Cobianchi, Karpel, Williams, Notario & Wilson, 1988; Kumar *et al.*, 1990; Casas-Finet *et al.*, 1993). Both the UP1 domain and the C-terminal domain are required for the alternative splicing activity of hnRNP A1, and within the UP1 domain, mutations in conserved surface residues of either RRM abolish activity, although general RNA binding is not substantially affected (Mayeda, Munroe, Cáceres & Krainer, 1994). Purified hnRNP A1 binds with high affinity to short RNA's containing one or more copies of the motif UAGGGA/U (Burd & Dreyfuss, 1994), although the physiological significance of this high-affinity binding remains controversial (Abdul-Manan, O'Malley & Williams, 1996). A 38-amino-acid domain near the C-terminus, termed M9, has been shown to be both necessary and sufficient for the nuclear localization of hnRNP A1 (Siomi & Dreyfuss, 1995). In addition, M9 is also a nuclear export signal (Michael, Choi & Dreyfuss, 1995), responsible for the shuttling of hnRNP A1 between the nucleus and the cytoplasm (Piñol-Roma & Dreyfuss, 1992).

The RRM is an ancient and extremely common RNA-binding motif (reviewed in Birney, Kumar & Krainer, 1993). Many proteins with RRM's have two, three, or four copies of the motif, and of these some, but not all, require more than one RRM for stable and specific RNA binding. The tertiary structures of several isolated RRM's have been recently determined, including the crystal and solution structure of the first of two RRM's in the U1 snRNP A polypeptide (Nagai, Oubridge, Jessen, Li & Evans, 1990; Hoffman, Query, Golden, White & Keene, 1991) and its co-crystal structure with bound RNA (Oubridge, Ito, Evans, Teo & Nagai, 1994), the NMR structure of the RRM of hnRNP C1/C2 with and without bound RNA (Wittekind, Görlach, Friedrichs, Dreyfuss & Mueller, 1992; Görlach, Wittekind, Beckman, Mueller & Dreyfuss, 1992) and the NMR structure of the second RRM of Sex-lethal (Lee, Kanaar, Rio & Wemmer, 1994). The secondary structure and folding topology of the first RRM of hnRNP A1 have also been solved by NMR (Garrett *et al.*, 1994). However, no structures with two or more RRM's have been determined to date. We hope that the three-dimensional structure of UP1 will not only facilitate the understanding of the structural basis of the RNA-binding specificity and alternative splicing activity of hnRNP A1, but also shed light on how synergy between RRM's may arise in other proteins with multiple RRM's. Here we report the crystallization and preliminary X-ray analysis of the UP1 domain of hnRNP A1.

2. Materials and methods

2.1. Overexpression and purification

The plasmid pET9d-UP1, which allows overexpression of the first 196 amino acids of human hnRNP A1 from a cDNA fragment cloned into the T7 expression vector pET-9d (Studier, Rosenberg, Dunn & Dubendorff, 1990), has been described (Mayeda *et al.*, 1994). BL21(DE3)pLysS *E. coli* cells carrying the pET9d-UP1 plasmid were grown in LB medium containing kanamycin ($100 \mu g \, ml^{-1}$) and chloramphenicol ($25 \mu g \, ml^{-1}$) at 310 K to an OD_{600} between 0.8 and 1.0, at which point isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of $0.4 \, mM$, and the temperature was reduced to 298 K. After 18 h, cells were harvested by centrifugation, washed, and stored at 203 K before use.

All purification steps were carried out at 277 K. The frozen cell pellet from a 2 l culture was thawed, resuspended in 40 ml of lysis buffer [$20 \, mM$ HEPES pH 7.5, $100 \, mM$ NaCl, 5% glycerol, $1 \, mM$ EDTA, $1 \, mM$ EGTA, $1 \, mM$ phenylmethylsulfonyl fluoride (PMSF) and 0.1% β-mercaptoethanol (BME)], and the cells were lysed by sonication. Cellular debris were removed by centrifugation and the cleared supernatant was brought to 50% ammonium sulfate saturation by slow addition of solid ammonium sulfate with gentle stirring. The precipitate was removed by centrifugation after 1 h incubation, and the cleared supernatant was diluted with lysis buffer to a final ammonium sulfate concentration of 40% saturation, prior to loading onto a 50 ml Phenyl–Sepharose column (Pharmacia). UP1 eluted from the column at approximately 12% saturated ammonium sulfate in a 40–0% linear gradient. Fractions containing UP1 were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE), pooled and dialyzed twice against 2 l of buffer *B* ($20 \, mM$ MES pH 6.0, 5% glycerol, $50 \, mM$ NaCl, $1 \, mM$ EDTA, 0.1% BME), and then loaded onto a 40 ml Heparin–agarose column (Bio-rad). UP1 eluted from the column at approximately $300 \, mM$ NaCl in a linear $50 \, mM$–$1 \, M$ gradient. Fractions containing UP1 were analyzed by SDS–PAGE, pooled and concentrated in Centricon (Amicon) concentrators to 15–$20 \, mg \, ml^{-1}$ for crystallization. Typically, a 2 l culture yielded approximately 30 mg of protein of sufficient purity for crystallization.

2.2. Crystallization and preparation of heavy-atom derivatives

Initial crystallization conditions were determined using the sparse-matrix screening method (Jancarik & Kim, 1991) by the hanging-drop vapor-diffusion technique (McPherson, 1982). After initial crystals were observed, the concentrations of precipitant, salt, and pH were carefully refined to reach optimal conditions. The best diffracting crystals were grown at 289 K in a $6 \mu l$ hanging drop containing equal volumes of protein solution ($15 \, mg \, ml^{-1}$ or higher in $300 \, mM$ NaCl, $20 \, mM$ MES, pH 6.0, $1 \, mM$ EDTA, 0.1% BME) and mother liquor [25% polyethylene glycol (PEG) 4000, 20% 2-methyl-2,4-pentanediol (MPD) or 10–15% glycerol, $100 \, mM$ Tris, pH 8.5]. Needle-like crystals appeared within a few hours after protein and precipitant were mixed. The crystals reached their maximum size of approximately $0.1 \times 0.1 \times 1.0 \, mm$ after 3 d (Fig. 1). Heavy-atom derivatives were prepared by transferring individual crystals into $6 \mu l$ mother liquor droplets on cover slips, and $3 \mu l$ of $5 \, mM$ mercurial or platinum compound solutions were introduced into the crystal-containing droplets. The crystals were soaked for 24–48 h before being subjected to the X-ray beam.

2.3. X-ray data collection

UP1 crystal diffraction data were collected at beamline X12C using an imaging-plate detector (MAR Research) at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL). UP1 crystals are susceptible to radiation damage at room temperature. Therefore, all data were collected at 100 K using an Oxford Cryosystems Cryostream cryocooling system. Individual crystals were picked up in nylon cryo-loops and directly dipped into liquid nitrogen before mounting the loops onto the goniometer head under a 100 K gaseous nitrogen stream. Images were indexed, integrated and scaled using the *HKL* suite of programs (Otwinowski, 1993). The space-group determination was further aided by a pseudo-precession display program, *PRECESS*, of the *PHASES* program suite (Furey & Swaminathan, 1997).

For heavy-atom derivative data collection, derivatization of the crystals was monitored prior to data collection by measuring the X-ray fluorescence while scanning the wavelength around the L_{III} absorption edges of Hg and Pt, for crystals soaked with mercurial and platinum compound solutions, respectively. When collecting derivative data sets, the incident X-ray wavelength was adjusted to a position corresponding to the absorption peak for the heavy atom used in order to maximize the anomalous differences, which will be used for phasing.

3. Results and discussion

The UP1 crystals belong to space group $P2_1$ with cell dimensions $a = 37.94$, $b = 43.98$, $c = 55.64 \, \text{Å}$ and $\beta = 93.9°$. From the predicted molecular mass of 22 264 Da and the crystal unit-cell volume, we determined that there is one molecule per asymmetric unit with a V_M of $2.1 \, \text{Å}^3 \, Da^{-1}$, and a calculated solvent content of 49%. The solvent content was obtained by the method described by
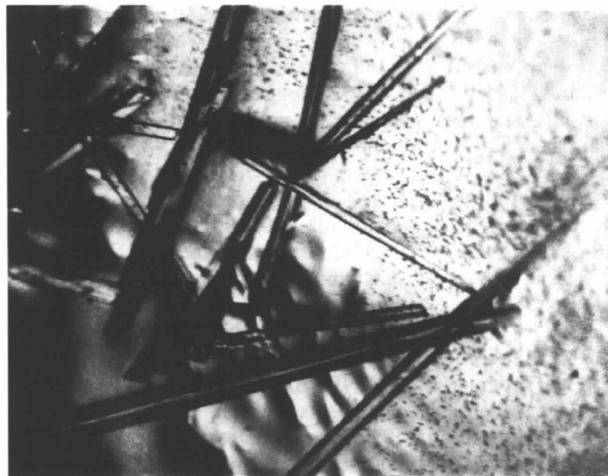

Fig. 1. Crystals of recombinant human UP1 protein. The largest crystal is approximately 0.1 mm in width and 1 mm in length.

Matthews (1985), using a partial specific volume of $0.762 \, \mathrm{ml \, g^{-1}}$ calculated from the amino-acid composition (Perkins, 1986) and the anhydrous crystallographic residue volumes derived from intact proteins (Chothia, 1975).

UP1 crystals diffracted X-rays strongly. At least in one case, diffraction beyond 1.3 Å resolution was observed (Fig. 2). Unfortunately, a high-resolution data set from this crystal was not collected because of the presence of satellite spots when it was rotated away from the initial orientation. Instead, a 2.0 Å data set was collected from a second crystal. This data set is 98.1% complete ($I/\sigma > 0$, 86% complete in the highest resolution shell of 2.03–2.00 Å) with an overall R_{merge} of 5.2%.

To solve the phase problem, we first tried the molecular replacement method using the crystal structure of the N-terminal RRM of U1A (Nagai *et al.*, 1990) as the search model. However, no significant rotational and translational solutions were found. Next, we soaked the crystals in $1.7 \, \mathrm{m}M \, \mathrm{K_2PtCl_4}$ solution for 48 h and collected a 2.4 Å data set with a 98.6% ($I/\sigma > 0$) completeness and an R_{merge} of 4.1%. This derivative showed one very strong peak (12σ) in the $V = 1/2$ section of the anomalous difference Patterson map. However, the isomorphous difference between the derivative and the native data set was significant only at a resolution lower than 4 Å. In the meantime, we have obtained two mercurial derivatives from *p*-chloromercuribenzene sulfonate (PCMBS) and methyl mercuric chloride (MMC) soaks. Both PCMBS and MMC derivatives have two metal sites, with nearly identical positions between the two derivatives. However, the relative occupancy between the two sites in the MMC derivative appears to be different from that in the PCMBS derivative. Therefore, the MMC derivative may contribute additional phasing information even though it has redundant metal positions. A 2.5 Å

PCMBS derivative data set was collected at $\lambda = 1.0$ Å with a 96.3% completeness ($I/\sigma > 1.0$) and an R_{merge} of 8.6%. Similarly, a 2.5 Å MMC derivative data set was collected at $\lambda = 1.0078$ Å with a 98.4% completeness ($I/\sigma > 1.0$) and an R_{merge} of 5.9%.
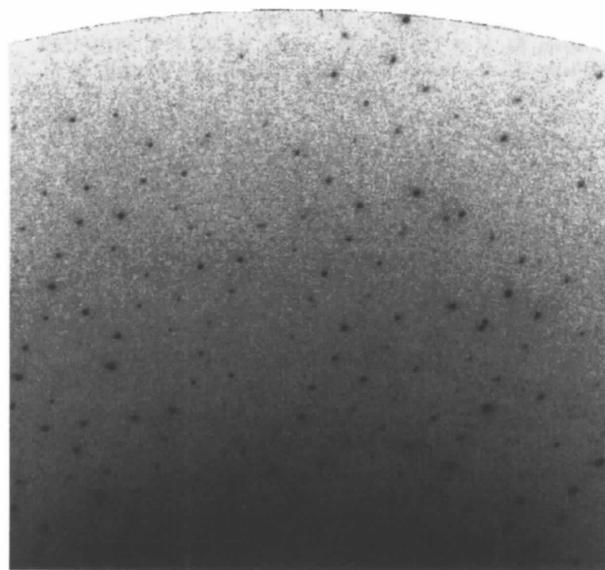

Fig. 2. A section of a 1.5° oscillation diffraction pattern collected at beamline X12C of NSLS, BNL, recorded with a 30 cm diameter MAR imaging-plate detector at a distance of 145 mm, $\lambda = 1.0$ Å and 90 s exposure time. The resolution at the edge of the photograph corresponds to 1.28 Å.
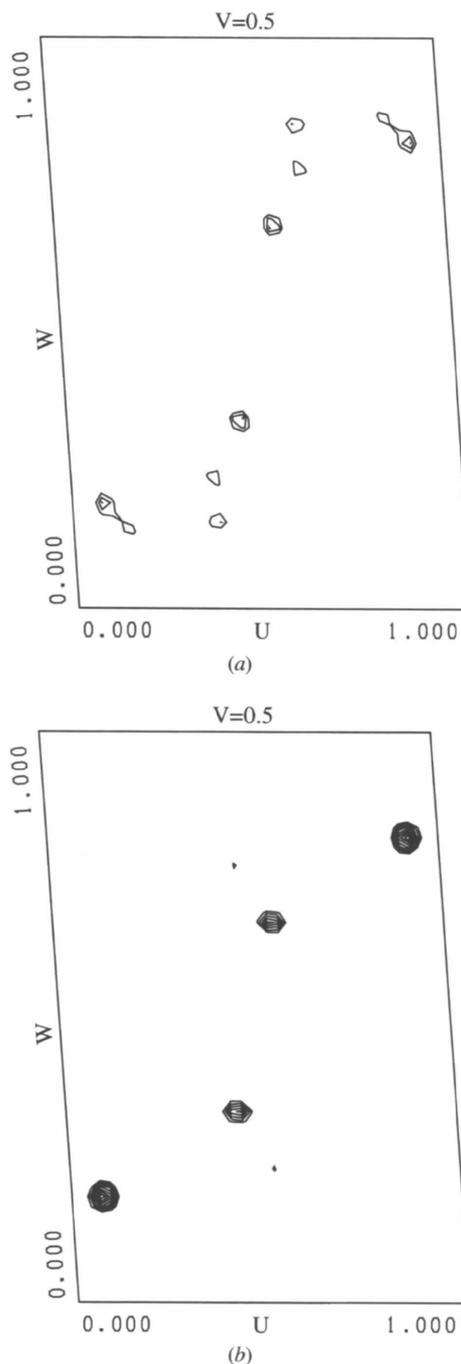

Fig. 3. (a) $V = 1/2$ section of the PCMBS isomorphous difference map at 2.5 Å resolution, contoured at 1σ interval starting at 3σ. (b) The same section for the anomalous difference map at 2.5 Å resolution with same contouring.

In summary, we have crystallized the recombinant human UP1 protein fragment. UP1 crystals diffract X-rays strongly, and two mercurial isomorphous derivatives have been obtained. These diffraction data should enable us to solve the phase problem for structure determination, which is currently in progress and will be reported elsewhere.

References

Abdul-Manan, N., O'Malley, S. M. & Williams, K. R. (1996). *Biochemistry*, **35**, 3545–3554.

Birney, E., Kumar, S. & Krainer, A. R. (1993). *Nucleic Acids Res.* **21**, 5803–5816.

Burd, C. G. & Dreyfuss, G. (1994). *EMBO J.* **13**, 1197–1204.

Buvoli, M., Cobianchi, F. & Riva, S. (1992). *Nucleic Acids Res.* **20**, 5017–5025.

Casas-Finet, J. R., Smith, J. D. Jr, Kumar, A., Kim, J. G., Wilson, S. H. & Karpel, R. L. (1993). *J. Mol. Biol.* **229**, 873–889.

Chothia, C. (1975). *Nature (London)*, **254**, 304–308.

Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V. & Wilson, S. H. (1988). *J. Biol. Chem.* **263**, 1063–1071.

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S. & Burd, C. G. (1993). *Ann. Rev. Biochem.* **62**, 289–321.

Fu, X.-D., Mayeda, A., Maniatis, T. & Krainer, A. R. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.

Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* In the press.

Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M. & Gronenborn, A. M. (1994). *Biochemistry*, **33**, 2852–2858.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. (1992). *EMBO J.* **11**, 3289–3295.

Herrick, G. & Alberts, B. (1976). *J. Biol. Chem.* **251**, 2124–2132.

Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W. & Keene, J. D. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.

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

Kiledjian, M. & Dreyfuss, G. (1992). *EMBO J.* **11**, 2655–2664.

Kumar, A., Casas-Finet, J. R., Luneau, C. J., Karpel, R. L., Merrill, B. M., Williams, K. R. & Wilson, S. H. (1990). *J. Biol. Chem.* **265**, 17094–17100.

Kumar, A. & Wilson, S. H. (1990). *Biochemistry*, **29**, 10717–10722.

Lee, A. L., Kanaar, R., Rio, D. C. & Wemmer, D. E. (1994). *Biochemistry*, **33**, 13775–13786.

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: John Wiley.

Matthews, B. W. (1985). *Methods Enzymol.* **114**, 176–187.

Mayeda, A., Helfman, D. M. & Krainer, A. R. (1993). *Mol. Cell. Biol.* **13**, 2993–3001.

Mayeda, A. & Krainer, A. R. (1992). *Cell*, **68**, 365–375.

Mayeda, A., Munroe, S. H., Cáceres, J. F. & Krainer, A. R. (1994). *EMBO J.* **13**, 5483–5495.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H. & Williams, K. R. (1988). *J. Biol. Chem.* **263**, 3307–3313.

Michael, W. M., Choi, M. & Dreyfuss, G. (1995). *Cell*, **83**, 415–422.

Munroe, S. H. & Dong, X. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 895–899.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). *Nature (London)*, **346**, 515–520.

Perkins, S. J. (1986). *Eur. J. Biochem.* **157**, 169–180.

Piñol-Roma, S. & Dreyfuss, G. (1992). *Nature (London)*, **355**, 730–732.

Pontius, B. W. & Berg, P. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 8403–8407.

Portman, D. S. & Dreyfuss, G. (1994). *EMBO J.* **13**, 213–221.

Otwinowski, Z. (1993). *Data Collection and Processing. Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Issacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

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

Siomi, H. & Dreyfuss, G. (1995). *J. Cell Biol.* **129**, 551–560.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.

Swanson, M. S. & Dreyfuss, G. (1988). *EMBO J.* **7**, 3519–3529.

Williams, K. R., Stone, K. R., LoPresti, M. B., Merrill, B. M. & Planck, S. R. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 5666–5670.

Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. (1992). *Biochemistry*, **31**, 6254–6265.