Interpreting a Low Resolution Map of Human U1 snRNP Using Anomalous Scatters

Chris Oubridge,1 Daniel A. Pomeranz Krummel,1,2 Adelaine K.-W. Leung,1,2 Jade Li,1 and Kiyoshi Nagai1,4

1Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, England, UK
2Present address: Department of Biochemistry, Brandeis University, Waltham, MA 02454, USA
3Present address: Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
4Correspondence: kn@rnc-lmb.cam.ac.uk
DOI 10.1016/j.str.2009.05.009
Open access under CC BY license.

SUMMARY

We recently determined the crystal structure of the functional core of human U1 snRNP, consisting of nine proteins and one RNA, based on a 5.5 Å resolution electron density map. At 5–7 Å resolution, α helices and β sheets appear as rods and slabs, respectively, hence it is not possible to determine protein fold de novo. Using inverse beam geometry, accurate anomalous signals were obtained from weakly diffracting and radiation sensitive P1 crystals. We were able to locate anomalous scatterers with positional errors below 2 Å. This enabled us not only to place protein domains of known structure accurately into the map but also to trace an extended polypeptide chain, of previously undetermined structure, using selenomethionine derivatives of single methionine mutants spaced along the sequence. This method of Se-Met scanning, in combination with structure prediction, is a powerful tool for building a protein of unknown fold into a low resolution electron density map.

INTRODUCTION

Most proteins in eukaryotic cells exist as components of large protein, RNA-protein, or DNA-protein complexes, which carry out important biological functions in an integrated manner (Gavin et al., 2002). The structure of individual components of these complexes often provides little insight into the structure and mechanism of the assembly of which they are a part. Therefore, the structure of all, or a major part of the assembly must be determined. For example, the structure of individual ribosomal proteins provided little or no insight into protein synthesis or decoding (Ramakrishnan and White, 1998). Therefore, to understand higher order functions of the cell it is important to undertake crystallographic studies of large macromolecular assemblies (Ramakrishnan, 2002; Klein et al., 2004). However, it is very difficult to purify many of these assemblies from natural sources or to assemble them from recombinant components in quantities sufficient for crystallographic studies (Maier et al., 2006; Liu et al., 2006). Crystallization is particularly challenging for large complexes because of sample heterogeneity, which may arise from modification of the sample in vivo, and dissociation of any weakly associated components of the complex as observed with RNA polymerase II (Cramer et al., 2000) and the ribosome (Wimberly et al., 2000). Furthermore, even when these difficulties are overcome, crystals of large complexes often diffract weakly to low resolution and are very susceptible to radiation damage. Fortunately, even low resolution diffraction data (5–7 Å) can provide essential insights into the functions of large macromolecular assemblies, provided one can obtain and interpret an electron density map (Ban et al., 1999; Clemons et al., 1999; Murakami et al., 2002; Bushnell et al., 2004). At such resolutions, α helices appear as tubular density (Muirhead and Perutz, 1963) and β sheets as flat density (Ban et al., 1999; Clemons et al., 1999). It may be possible to place a protein of known structure into the electron density map at this resolution, particularly if its secondary structure is largely α-helical; it is, however, difficult to follow the polypeptide chain of a protein of unknown structure and determine its fold de novo.

We have recently solved the structure, at 5.5 Å resolution, of the functional core of human U1 snRNP, which consists of U1 snRNA, seven Sm proteins, U1-70K, and U1-C (Pomeranz Krummel et al., 2009). In this paper, we show that anomalous peaks from selenium and other heavy atoms could be obtained from highly radiation-sensitive crystals that typically diffracted to ~6.5 Å. The anomalous peaks were an important aid to fitting known protein folds into the electron density and provided evidence of the quality of the fit. A polypeptide chain of unknown structure was traced on the basis of selenium positions when methionines, which could subsequently be replaced by selenomethionine (SeMet), were introduced by mutagenesis into the sequence at short intervals.

RESULTS

Calculation of an Experimental Electron Density Map at 6.5 Å Resolution

The reconstitution and crystallization of U1 snRNP have been described (Muto et al., 2004; Pomeranz Krummel et al., 2009). The functional core of human U1 snRNP consists of U1 snRNA and nine proteins. The crystals grow in P1 space group with unit cell parameters a = 127 Å, b = 128 Å, c = 156 Å, α = 96°, β = 107°, and γ = 101° and diffract to ~6 Å resolution. Self-rotation and self-Patterson analyses suggested four U1 snRNPs in the asymmetric unit (ASU) (data not shown).

A multiwavelength anomalous dispersion data set was collected from a tantalum bromide cluster (TaBr12) derivative (Knäblein et al., 1997) at the Ta L-III edge at two wavelengths:
The inflection data were used to calculate an anomalous Patterson map (Figure 1A) and the coordinates of four $Ta_6Br_{12}$ sites were obtained manually from the cross-peaks. Ta$6Br_{12}$ cluster coordinates and occupancies were refined in SHARP (de la Fortelle and Bricogne, 1997). Inspection of residual maps showed four additional minor sites with lower occupancy.

Each minor site was 48 Å from a major site, confirming that there were four U1 snRNPs in the ASU, related by noncrystallographic symmetry (NCS), and each bound to two Ta$6Br_{12}$ clusters. Spherically averaged form factors of the clusters at 7 Å resolution resulted in higher final phasing power (1.51 versus 1.25), lower Cullis R factor (0.71 versus 0.76), and better overall figures of merit (0.413 versus 0.404) than a single point Gaussian model.

Figure 1B shows the packing of four U1 snRNPs in the unit cell and the positions of the four major and four minor Ta sites. The sites were refined with and without coordinate inversion, and the phases were subjected to solvent flipping in Solomon (Abrahams and Leslie, 1996) with a 60% solvent content and extended from 7.5 to 7.0 Å over 11 cycles. The correct hand was identified from better figures of merit for the solvent flattened phases (0.541 versus 0.531) and clear density for A-form RNA in the resulting electron density map.

To improve phasing and to aid location of the Sm proteins, they were all labeled with SeMet then incorporated into U1 snRNP crystals. Se K-edge ($\lambda = 0.9797$ Å) data were collected and an anomalous difference map calculated using solvent-flattened Ta$6Br_{12}$ phases. The map had 123 peaks greater than 3.0 SD ($\sigma$) above background (data not shown).

To further improve phasing and to locate the Zn-finger protein U1-C, an anomalous map was calculated from native data collected at the zinc K edge ($\lambda = 1.2827$ Å). The map showed...
Interpreting a Low Resolution Map of U1 snRNP

Structure

Figure 2. Overlay of Selenium Peaks from Multiple Crystals Containing U1-70K SeMet Mutant Protein

(A) U1-70K residues 61–180 are shown as orange cartoon, with part of U1 snRNA, including SL1, shown in light gray. The selenium peak coordinates from anomalous maps of the eight U1-70K mutants are marked by colored spheres. The selenium anomalous maps are shown, all contoured at 3.5 σ and colored to match the spheres. Sphere diameter is ~2 Å. As well as the four natural methionines (67, 88, 134, and 157), which have corresponding peaks in all the mutants, two of the mutant site peaks (E61M and I75M) are also shown. The colors are: wild-type, black; L9M, red; I19M, orange; E31M, light green; I41M, dark green; I49M, cyan; E61M, blue; I75M, dark blue. (B) The path of the extended N terminus of U1-70K. Electron density attributed to U1-70K is shown in brown and contoured at 1 Å. Where density is absent, approximately between residues 24 and 45, a plausible path for the peptide is indicated based on the selenium positions of E31M and I41M. Selenium peaks and anomalous maps are as for (A). Near the selenium site of L9M, U1-70K is seen to interact with U1-C, which is red.

Model Building

RNA

The resulting electron density map was of high quality such that RNA helices, with shallow minor grooves and deep major grooves, were readily discernible. It was not always possible to fit long fragments of A-form RNA into extended regions of helical density, they were therefore initially built from short fragments of idealized A-form RNA and manually rebuilt around the junctions to improve stereochemistry. An NMR structure of a kissing loop RNA (Kim and Tinoco, 2000; PDB code 1f5u) was built into the electron density at the apical part of stem loop 2 (SL2). This domain had been introduced into the RNA to promote crystal lattice interactions (Pomeranz Krummel et al., 2009), which are seen within the ASU between U1 snRNP complexes A and B and complexes C and D (Figure 1B).

Sm Proteins

The structure of the U4 snRNP core domain, an assembly of seven Sm proteins and a 68 nucleotide RNA, was solved to 3.6 Å resolution independently (A.K.W.L., J.L., and K.N., unpublished data) and four of the protein structures have been determined as heterodimers of SmD1D2 and SmBD3 (Kambach et al., 1999). One to one correspondence was found between the majority of our selenium anomalous peaks from SeMet-labeled Sm proteins and the methionine residues in the U4 core domain whereas some anomalous peaks are composite peaks arising from two or more selenium atoms. The U4 core domain structure (A.K.W.L., J.L., and K.N., unpublished data) was initially fitted into the U1 snRNP map by superposition of the U4 core methionine sulfur positions onto the corresponding refined Se sites. In the U1 snRNP map the N-terminal helices of the Sm fold were fitted into tubular densities but the electron density for the β sheet in some subunits was discontinuous.

U1-C

The lowest energy NMR structure of the Zn-finger domain of U1-C (Muto et al., 2004; PDB code 1uw2) was placed at the Zn anomalous peak positions based on the rod-like density of the α-helical region. A long section of α-helical density from U1 snRNP complex A extends toward SL3 of complex C and vice versa, and complexes B and D show the same relationship. This long α-helix of U1-C shows that helices B and C, which form a turn with a flexible loop in isolated U1-C in the solution structure (Muto et al., 2004), form a continuous helix in the U1 snRNP crystal. In order to ensure correct orientation of the Zn-finger domain a native crystal was treated with ethyl mercury thiosalicylate (EMTS). An anomalous peak of mercury (Hg) bound to Cys-25 adjacent to Zn-coordinating His-24 was used to place the Zn-finger domain more precisely. An additional Hg anomalous peak from a single Cys mutant (Q39C) was used to orient the C-terminal helix and place it in the correct register.

U1-70K

The RNA binding domain (RBD) of U1-70K was known to bind to SL1 (Patton and Pederson, 1988; Query et al., 1989) and a large globule of electron density was seen in the SL1 loop region. The initial map did not permit unambiguous fitting of the RBD, so U1-70K was labeled with SeMet and reconstituted into U1.

four peaks greater than 7.8 σ with the next strongest being 4.4 σ. The four strongest peaks correspond to zinc atoms in U1-C protein, in agreement with there being four U1 snRNP complexes in the ASU.

The coordinates of Se atoms, Zn atoms, and Ta6Br12 clusters were refined together against their respective data sets in SHARP (de la Fortelle and Bricogne, 1997). Occupancy and position were refined for all sites, except for one Ta6Br12 site, which was fixed at the origin. B factors were fixed at 200.0 Å². Occupancy and B factors were refined together against their respective data sets in SHARP (de la Fortelle and Bricogne, 1997). Occupancy and position were refined for all sites, except for one Ta6Br12 site, which was fixed at the origin. B factors were fixed at 200.0 Å². The phases were improved by solvent flipping in Solomon with the next strongest being 4.4 σ. The coordinates of Se atoms, Zn atoms, and Ta6Br12 clusters were refined together against their respective data sets in SHARP (de la Fortelle and Bricogne, 1997). Occupancy and position were refined for all sites, except for one Ta6Br12 site, which was fixed at the origin. B factors were fixed at 200.0 Å².
snRNP. The resulting crystal gave four Se anomalous peaks (above 4.0 $\sigma$) per U1 particle. Two of the peaks, corresponding to Met-134 and Met-157, lie within the RBD (Figure 2A). The RBD was homology modeled from the N-terminal RBD of U1A (Nagai et al., 1990; Oubridge et al., 1994) and placed using these peaks along with the rod-like density of its two $\alpha$ helices. The loop region of SL1 was built in such a way that G28 and U30 are in close proximities of Tyr112 and Leu175, which were assumed from cross-linking data (Urlaub et al., 2000). The remaining two anomalous peaks were found in the long rod-like density adjacent to SL1. This region was predicted to be $\alpha$-helical and was modeled as such between residues 63 and 89. Further support for this model comes from the observation that many of the basic residues of the helix are close to the phosphate backbone of SL1, favoring electrostatic interactions (Pomeranz Krummel et al., 2009).

Figure 3. Superposition of NCS-Related Particles

The four particles in the ASU were superimposed in O (Jones et al., 1991) using the C$\alpha$ atoms of Sm protein residues within the Sm fold, U1-70K residues 9–23, and the P atoms of U1 snRNA nucleotides 1–16 and 48–134. Particles B, C, and D were transformed onto particle A with rms of 1.75 Å, 2.03 Å, and 1.97 Å for those C$\alpha$ and P atoms, respectively. The particles are shown as ribbons with A, B, C, and D in blue, green, yellow, and red, respectively. The substructures that varied in their relative orientations between particles, and which were treated as separate domains in averaging, are indicated. SL1, U1-70K RBD, and U1-70K helix (63–89) were combined and treated as a single domain. SL4 and U1-C C-terminal helix (32–61) were both treated as separate domains.

The location and the structure of the N-terminal 62 residues of U1-70K were unknown, so we made several single methionine mutants in this region to locate these residues from the position of anomalous peaks when U1 snRNP was reconstituted with their SeMet derivative. The amino acids chosen for substitution with methionine had long and/or hydrophobic side chains, in order to minimize structural changes. Basic side chains were avoided because their mutation could disrupt RNA binding. Substitutions were made at approximately ten residue intervals to facilitate chain tracing. Data from crystals grown with SeMet derivatives of these mutants gave selenium anomalous difference peaks that revealed an extended polypeptide from residue 9 to 63 (Figure 2B). The polypeptide interacts with U1-C protein and traces a path around the Sm protein ring to the beginning of the $\alpha$ helix adjacent to SL1. Figure 2A shows an overlay of Se peaks for the natural methionine residues of U1-70K that are present in all these crystals. The selenium anomalous peaks of these residues are clustered with rms < 1.9 Å, except for Met67 in complex B (rms = 2.5 Å). The observed scatter was within the expected positional errors at this resolution.

Phase Extension to 5.5 Å: Multi-NCS, Multi-Crystal Averaging

An EMTS-soaked crystal of U1 snRNP, containing the Q39C mutant of U1-C, diffracted to 5.5 Å, compared with 6.5 Å for the Sm protein SeMet derivative crystal from which, along with the $TaBr_{12}$ derivative and zinc edge data, the original experimental map was calculated. The unit cell of this crystal had a c axis 3.7 Å shorter than the mean of the other crystals’ c (152.0 Å compared to average of 155.7 ± 0.7 Å; see Table S1 in Pomeranz Krummel et al. (2009) for crystal statistics). The EMTS-soaked U1-C Q39C
were generated in O (Jones et al., 1991). Masks were created for each of these substructures using the program NCSMASK (CCP4, 1994). The masks, the solvent-flattened phases at 6.5 Å, the SeMet Sm core data, and the U1-C(Q39C)-EMTS crystal data were used for multidomain, multi-crystal averaging in the program DMMULTI (Cowtan et al., 2001). This resulted in phases with mean figures of merit of 0.623 to 5.5 Å, which compares well with a value of 0.652 to 6.5 Å for the phases used to calculate the experimental map. The resulting 5.5 Å map was clearly of higher quality than the original map: density, extending the N-terminal region from that seen in the U4 core structure. Many of these are connected to SmD2 by a kinked rod-like density suggesting a helical structure. This peak is attributed to Met11 of SmD2. If these helices are built into the density, extending the N-terminal region from the SmD1-D2 heterodimer (Kambach et al., 1999), then Met11 can account for the peak. The U1 map reveals other regions that have shifted relative to the U4 core structure. Many of these are the Salt and C termini, outside the canonical Sm fold. One of the most notable changes is for SmF protein residues 6 to 15. There are also conformational differences in some loop regions within the Sm fold, such as SmD3 residues 49 to 55 between strands I3 and I4 and SmF residues 46 to 56, which are explainable by interaction with the N-terminal peptide of U1-70K and with a neighboring complex, respectively. Overall, the structures appear to be the same within the canonical Sm fold.

Accuracy of U1-70K Selenium Positions at 5.5 Å Resolution

The U1-70K protein mutants provide useful information regarding the accuracy of atom positions determined by this method. We were able to trace the extended N-terminal region of U1-70K by SeMet labeling. What is the accuracy of selenium positions determined by this method? Table 1 summarizes the data statistics of the crystals containing these Se-Met-labeled proteins. The means and SD are given for each of the cell dimensions. The largest variations are seen with the c axis, which has an SD of 0.36 Å, the others have values of 0.12–0.17 Å.

All the mutants contain common natural methionine residues: two in the RBD (Met-134 and Met-157) and two in the α-helical region (Met-67 and Met-88). Figure 2A shows the positions of the anomalous peaks together with the model of the RBD and
The Se atoms of the mutant sites all gave significant anomalous peaks, except for the E49M site in particle C (Table 2). We also compared the positions of the mutant Se-Met sites by overlaying the U1-70K mutant and wild-type anomalous peak positions of the four particles. The mutant Se anomalous peaks all superimposed with SD values of less than 2 Å (Table 3).

**DISCUSSION**

We obtained our initial electron density map using a tantalum bromide derivative. Data were collected in inverse beam mode with 1° wedges to measure anomalous signals accurately (see Data Collection in Experimental Procedures). This strategy resulted in a strong anomalous signal. Surprisingly, when the wedge size was increased to 5°, we obtained a much weaker anomalous signal from several different crystals, as judged by correlation coefficients within (l+) and (l-), measured in SCALAL (CCP4, 1994; Evans, 2006), and the lack of significant peaks in the anomalous difference map. We attribute this to rapid decay of the anomalous signal during data collection. For the radiation-sensitive P1 crystal of U1 snRNP the use of inverse beam mode was absolutely essential. We showed that the position of heavy atom scatterers such as Se, Zn, and Hg could be

---

**Table 1. The Cell Parameters and Data Statistics of the U1 snRNP Crystals with the SeMet Derivatives of Different U1-70k Mutants**

| SeMet U1-70k Data set | Multiplicity | Anomalous Multiplicity | Completeness | Anomalous Completeness | Rmerge | Rp,i.m. | Mn(I)/sd(I) |
|-----------------------|--------------|------------------------|--------------|------------------------|--------|---------|-------------|
| Wild-type             | 13.0 (13.3)  | 6.5 (6.7)              | 88.2 (89.8)  | 86.7 (88.1)            | 21.4   | 6.8     | 43.7 (6.4)  |
| L9M                   | 4.2 (4.2)    | 2.1 (2.1)              | 98.7 (98.9)  | 98.3 (98.7)            | 4.5    | 3.0     | 15.5 (2.3)  |
| I19M                  | 9.9 (9.9)    | 4.9 (5.0)              | 99.1 (99.2)  | 98.9 (98.9)            | 6.7    | 2.4     | 9.3 (2.7)   |
| E31M                  | 7.8 (7.8)    | 3.9 (3.9)              | 98.6 (98.6)  | 98.2 (98.3)            | 10.3   | 4.2     | 15.3 (2.2)  |
| I41M                  | 8.3 (8.3)    | 4.1 (4.2)              | 98.7 (98.9)  | 98.4 (98.8)            | 6.3    | 2.5     | 17.9 (2.5)  |
| E49M                  | 8.3 (8.4)    | 4.2 (4.2)              | 99.1 (99.2)  | 98.7 (99.2)            | 11.4   | 4.4     | 15.5 (3.2)  |
| E61M                  | 8.3 (8.3)    | 4.1 (4.2)              | 98.7 (98.7)  | 98.4 (98.7)            | 7.4    | 2.9     | 16.3 (1.9)  |
| I75M                  | 8.3 (8.3)    | 4.1 (4.2)              | 98.8 (98.9)  | 98.4 (98.7)            | 11.1   | 4.3     | 14.9 (2.0)  |

**Table 2.** Most sites are above 4σ when the sites from the four particles are superimposed. Table from the mean of the anomalous peak positions for each site. The anomalous signal during data collection. For the radiation-sensitive P1 crystal of U1 snRNP the use of inverse beam mode was absolutely essential. We showed that the position of heavy atom scatterers such as Se, Zn, and Hg could be

---

`\[ R_{\text{merge}} = \frac{\sum_{hkl} \sum_{i} I_i(hkl) - \langle I(hkl) \rangle}{\sum_{hkl} \sum_{i} I_i(hkl)} \]`

`\[ R_{p,i.m.} = \frac{\sum_{hkl} \left[1/(N - 1)\right]^{1/2} \sum_{i} I_i(hkl) - \langle I(hkl) \rangle}{\sum_{hkl} \sum_{i} I_i(hkl)} \]`
determined with a positional error of <2 Å at 5.5 Å resolution (Tables 3 and 4). We were able to place the 3.6 Å structure of the U4 core domain by least-squares fit of methionine sulfur atoms onto anomalous peak positions in the U1 map with an rmsd of <2.4 Å. The agreement between the electron density and the U4 structure is excellent within the Sm fold but the map also indicates some differences between the structures in regions for which flexibility would be expected, such as loops and N- and C-terminal extensions beyond the Sm fold. It is not difficult to place a protein of known structure into an electron density map at this resolution and any ambiguities may be resolved by judicious use of heavy atom markers. We were able to trace the extended chain of the N terminus of U1-70K using Se-Met scanning. Hence, in principle it will be possible to trace the chains of other proteins of unknown structure with the same method at resolutions as low as 6.5 Å. Residues 63–88 of U1-70K were predicted to form an α helix using Jpred3 (Cole et al., 2008). When an idealized α helix was fitted into a tubular density, the distance between the two methionines (Met-67 and Met-88) and their relative orientations with respect to the helical axis were consistent with the model. With Se-Met

| Site       | Particle in ASU | Number of Atoms | SD from Mean (Å) | Mean SD for Site |
|------------|----------------|-----------------|------------------|-----------------|
| Se-67      | A              | 5               | 1.7              | 1.9             |
|           | B              | 7               | 2.7              |                 |
|           | C              | 7               | 1.9              |                 |
|           | D              | 8               | 1.4              |                 |
| Se-88      | A              | 7               | 1.1              | 1.6             |
|           | B              | 7               | 1.7              |                 |
|           | C              | 8               | 1.9              |                 |
|           | D              | 7               | 1.6              |                 |
| Se-134     | A              | 8               | 1.1              | 1.0             |
|           | B              | 8               | 1.0              |                 |
|           | C              | 8               | 1.0              |                 |
|           | D              | 8               | 1.0              |                 |
| Se-157     | A              | 8               | 1.3              | 1.2             |
|           | B              | 8               | 1.2              |                 |
|           | C              | 8               | 1.1              |                 |
|           | D              | 7               | 1.1              |                 |

Three matrices were generated for the superposition of U1-70k selenium sites from U1 snRNP complexes B, C, and D onto the U1-70k selenium sites of complex A, using the program O (Jones et al., 1991). An SD from the mean was then calculated for each site from the transformed coordinates.

\[
SD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2}
\]

was mutated to methionine, its position and orientation were also consistent with the model. This shows that α helices can be placed at correct position and orientation with a positional error of 2 Å at this resolution, providing important information that can be tested experimentally. In contrast with α helices, extended polypeptide strands are not always resolved at this resolution. Introduction of Se-Met at key positions provided landmarks to reveal the path of a strand of polypeptide, even

| Residue/Site | Number of Atoms | SD from the Mean (Å) |
|--------------|-----------------|----------------------|
| L9M          | 4               | 1.4                  |
| I19M         | 4               | 1.2                  |
| E31M         | 4               | 1.3                  |
| I41M         | 4               | 1.9                  |
| I49M         | 3               | 1.9                  |
| E61M         | 4               | 1.4                  |
| I75M         | 4               | 1.6                  |
| M67          | 27              | 2.5                  |
| M88          | 29              | 1.6                  |
| M134         | 32              | 1.1                  |
| M157         | 31              | 1.2                  |
where protein density was not observed in the experimental maps. In combination with secondary structure prediction, this method can elucidate more significant structural information from diffraction data at moderate resolution than can the electron density alone.

**EXPERIMENTAL PROCEDURES**

**SeMet Derivatives**
All SeMet-labeled proteins were expressed in a methionine auxotrophic strain of *E. coli* using the method as described in Ramakrishnan et al. (1993). U1-70K was expressed as an N-terminal fusion with thioredoxin followed by a (His) tag and a TEV protease cleavage site (Ignjatovic, 1999). The protein was purified on a Ni-NTA agarose column followed by proteolysis with TEV protease and a second Ni-NTA agarose column. Sm proteins were purified as previously described (Kambach et al., 1999). Reconstitution and crystallization of U1 snRNP have been described previously (Muto et al., 2004; Pomeranz Krummel et al., 2009).

**Heavy Atom Derivatization**
A tantalum bromide cluster (TaBr5; Knäbelein et al., 1997) derivative was prepared by directly adding a few grains of the dark green solid into crystallization drops and incubating for 6–8 hr at 4°C. The grains dissolved and the cluster compound soaked into the crystals. After incubation the crystals turned a darker shade of green than the surrounding mother liquor and were cryo-cooled as described below, without back soaking. Crystals were grown by hanging drop vapor diffusion from U1 snRNP reconstituted with the Q39C mutant of U1-C(1-77) protein (Pomeranz Krummel et al., 2009). The crystal mother liquor was exchanged with EMTS dissolved in the well buffer to a final concentration of 1 mM and incubated at 4°C for 20 hr. The crystal was back soaked in well buffer and cryo-cooled as described below.

**Cryo-Cooling**
Crystals were grown in the presence of 38%–40% 2-methyl-2,4-pentanediol and did not require further cryo-protection. Cryo-cooling was carried out by lifting crystals from the mother liquor on loops at 4°C and flash cooling in the gas jet of a cryostream (Oxford Cryosystems) at 100 K. All these steps were done at 4°C. Freezing in a cryostream consistently reduced the mosaicity of the crystals, as opposed to plunging the crystal directly into liquid nitrogen.

**Data Collection**
Diffraction data were collected on a Mar225 CCD detector at Swiss Light Source X06SA and X10SA beam lines with the X-rays focused on the detector. In order to minimize the effect of radiation damage inverse beam mode was used; i.e., after every 1° oscillation the crystal was rotated 180° such that Friel-del pairs were measured nearly simultaneously. Data were integrated with MOSFLM (Leslie, 2006) and scaled with SCALA (CCP4, 1994; Evans, 2006). Phasing has been described in the text as well as in Pomeranz Krummel et al. (2009).

Figures were prepared using PyMol (DeLano, 2002) except for Figure 1A, which was made in Mapslicer (CCP4, 1994).

**SUPPLEMENTAL DATA**
Supplemental Data include one movie and can be found with this article online at http://www.cell.com/structure/supplemental/S0969-2126(09)00220-2.

**ACKNOWLEDGMENTS**
We thank the staff at Daresbury, European Synchrotron Radiation Facility, and Swiss Light Source beamlines, particularly Clemens Schulze-Briese, Takashi Tomizaki, and Anuschka Pauluhn for their dedication during data collection; Phil Evans, Andrew Leslie, Harry Powell, and Clemens Vornhein for advice about crystallography; Jan Löwe for a gift of TaBr5; Venki Ramakrishnan for helpful discussion and support; and the past and present members of the Nagai group for innumerable contributions. The project was funded by the Medical Research Council and Human Frontier Science Program. D.A.P.K. was a recipient of a Human Frontier Science Program long-term fellowship. A.K.W.L. was supported by the National Science and Engineering Research Council of Canada, the ORS Fund, the Cambridge Commonwealth Trust, and Sidney Sussex College Junior Research Fellowship.

Received: August 27, 2008
Revised: April 21, 2009
Accepted: May 7, 2009
Published: July 14, 2009

**REFERENCES**
Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D Biol. Crystallogr. 52, 30–42.
Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B., and Stitz, T.A. (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. Nature 400, 841–847.
Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Ångstroms. Science 303, 983–988.
Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763.
Clemons, W.M., Jr., May, J.L., Wimberly, B.T., McCutcheon, J.P., Capel, M.S., and Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. Nature 400, 833–840.
Cole, C., Barber, J.D., and Benton, G.J. (2008). The Jpred3 secondary structure prediction server. Nucleic Acids Res. 36, W197–W201.
Cowtan, K.D., Zhang, K.Y.J., and Main, P. (2001). DM/MMULTI software for phase improvement by density modification. In International Tables for Crystallography, Volume F. Crystallography of Biological Macromolecules, M.G. Rossmann and E. Arnold, eds. (Dordrecht, Netherlands: Kluwer Academic Publishers), pp. 705–710.
Cramer, P., Bushnell, D.A., Fu, J., Gnant, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 289, 640–649.
de la Fortelle, E., and Bricogne, G. (1997). Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiewavelength anomalous diffraction methods. In Methods in Enzymology, C.W. Carter and R.M. Sweet, eds. (New York: Academic Press), pp. 472–493.
DeLano, W.L. (2002). The PyMOL Molecular Graphics System. (http://www.pymol.org).
Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. 62, 72–82.
Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.
Ignjatovic, T. (1999). Overproduction and characterisation of human U1 70k protein. Part II Dissertation, (Cambridge, UK: Cambridge University).
Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47, 110–119.
Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96, 375–387.
Kim, C.H., and Tinoco, I., Jr. (2000). A retroviral RNA kissing complex containing only two G-C base pairs. Proc. Natl. Acad. Sci. USA 97, 9396–9401.
Klein, D.J., Moore, P.B., and Steitz, T.A. (2004). The roles of ribosomal proteins in the structure, assembly and evolution of the large ribosomal subunit. J. Mol. Biol. 340, 141–177.

Knäblein, J., Neuefeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B., and Huber, R. (1997). Ta_{6}Br_{2+}, a tool for phase determination of large biological assemblies by X-ray crystallography. J. Mol. Biol. 270, 1–7.

Leslie, A.G.W. (2006). The integration of macromolecular diffraction data. Acta Crystallogr. D Biol. Crystallogr. 62, 48–57.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell 127, 1223–1237.

Maier, T., Jenni, S., and Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 Å resolution. Science 317, 1258–1262.

Muirhead, H., and Perutz, M.F. (1963). Structure of haemoglobin. A three-dimensional Fourier synthesis of reduced human haemoglobin at 5.5 Å resolution. Nature 199, 633–638.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science 296, 1285–1290.

Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D., and Nagai, K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. J. Mol. Biol. 341, 185–198.

Nagai, K., Oubridge, C., Jessen, T.-H., Li, J., and Evans, P.R. (1990). The crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature 348, 515–520.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature 372, 432–438.

Patton, J.R., and Pederson, T. (1988). The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. Proc. Natl. Acad. Sci. USA 85, 747–751.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. Nature 458, 475–480.

Query, C.C., Bentley, R.C., and Keene, J.D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70k U1 snRNP protein. Cell 57, 89–101.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell 108, 557–572.

Ramakrishnan, V., and White, S.W. (1998). Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. Trends Biochem. Sci. 23, 208–212.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature 362, 219–223.

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Lührmann, R. (2000). A general approach for identification of RNA-protein cross-linking sites within native human spliceosomai small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. J. Biol. Chem. 275, 41458–41468.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vornhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 3OS ribosomal subunit. Nature 407, 327–339.