crystallization papers

Crystallization, microPIXE and preliminary crystallographic analysis of the complex between the third KH domain of hnRNP K and single-stranded DNA

Paul H. Backe,[a] Raimond B. G. Ravelli,[a] Elspeth Garman[b] and Stephen Cusack[a]*

[a]European Molecular Biology Laboratory, Grenoble Outstation, 6 Rue Jules Horowitz, BP 181, F-38042 Grenoble CEDEX 9, France, and [b]Laboratory of Molecular Biophysics, Department of Biochemistry, Oxford University, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

Correspondence e-mail:
cusack@embl-grenoble.fr

hnRNP K is one of the major proteins found in hnRNP particles which are ribonucleoprotein complexes containing proteins and pre-mRNA. hnRNP K contains hnRNP K homology (KH) domains which bind both CT-rich single-stranded DNA (ssDNA) and CU-rich ssRNA. Co-crystallization of the third KH domain of human hnRNP K with a 15-mer ssDNA gave rod-shaped crystals belonging to the trigonal space group $P3_121$ (unit-cell parameters $a = 54.0$, $c = 149.7$ Å) and diffracting to 2.4 Å resolution. MicroPIXE (proton-induced X-ray emission) experiments showed that the crystals contained the complex and that the phosphorus to sulfur atomic ratio was consistent with the asymmetric unit containing three KH3 domains per 15-mer ssDNA. This was confirmed by structure solution by molecular replacement.

1. Introduction

As soon as pre-mRNA emerges from the transcription machinery, the nascent transcripts are associated with a group of specific RNA-binding proteins known as heterogeneous nuclear ribonucleoproteins (hnRNP proteins). These hnRNP proteins are involved in pre-mRNA processing and are important determinants of mRNA export, localization, translation and stability (Dreyfuss *et al.*, 2002). At least 20 major hnRNP proteins have been identified and are designated A1 (34 kDa) to U (120 kDa) (Dreyfuss *et al.*, 1993). hnRNP K is one of the major proteins found in hnRNP complexes. It gives its name to the hnRNP K homology (KH) domain, a compact single-stranded RNA- and DNA-binding module, of which there are three in hnRNP K. hnRNP K belongs to the family of poly-C-binding proteins (Makeyev & Liebhaber, 2002), but is unusual in binding to both CU-rich ssRNA (Ostareck *et al.*, 1997) and CT-rich ssDNA (Takimoto *et al.*, 1993). The protein contains both a classical nuclear-localization signal and a novel hnRNP K nuclear-shuttling (KNS) domain that allows the protein to shuttle between the nucleus and the cytoplasm (Michael *et al.*, 1997). Numerous studies have shown that hnRNP K has diverse specific functions (for reviews, see Bomsztyk *et al.*, 1997; Ostareck-Lederer *et al.*, 1998). In the nucleus, these include transcriptional activation of the human c-*myc* gene by binding specifically to a C-rich DNA sequence, termed the CT element, within the c-*myc* promotor (Michelotti *et al.*, 1996; Takimoto *et al.*, 1993; Tomonaga & Levens, 1996). In the cytoplasm, hnRNP K functions to regulate the translation of erythroid 15-lipoxygenase (LOX) by

binding to the differentiation-control element (DICE) on the 3′ untranslated region (UTR) of LOX mRNA (Ostareck *et al.*, 1997). hnRNP K has now been implicated in the regulation of gene expression at many levels, including transcription, mRNA processing, mRNA transport, mRNA stability and translation. hnRNP K itself seems to be regulated by phosphorylation, *e.g.* by c-Src kinase (Ostareck-Lederer *et al.*, 2002). Such a broad involvement suggests that it acts as a bridging or platform molecule that facilitates the association of proteins in multi-component regulatory and signalling complexes (Ostareck-Lederer *et al.*, 1998).

In this paper, we report the crystallization, microPIXE and preliminary X-ray diffraction data analyses of the third KH domain of hnRNP K in complex with an ssDNA derived from the c-*myc* promotor CT element. The crystal structure of such a complex will help to further elucidate the nucleic acid-binding properties of KH domains.

2. Materials and methods

2.1. Crystallization

The third KH domain (KH3) of human hnRNP K was expressed and purified as described previously (Backe *et al.*, 2004). The 15-nucleotide ssDNA oligonucleotide 5′-TTC CCC TCC CCA TTT (MWG Biotech AG) used here was identical to that previously studied by NMR (Baber *et al.*, 2000). Pure KH3 was concentrated to about 40 mg ml^{-1} (Ultra-free-15 BioMax 5K centrifugal filter device) as determined by the UV absorbance using a theoretical ε_{280} value of 2560 M^{-1} cm^{-1} and mixed with an equimolar amount of the

ssDNA. The mixture was then incubated on ice for approximately 30 min and spun at 2000g for 10 min prior to crystallization trials. Initial screening for crystallization conditions was performed using sparse-matrix screens (Jancarik & Kim, 1991) from Hampton Research. All crystallization experiments were performed at 294 K by the hanging-drop vapour-diffusion method in Linbro plates. Small fragile needles were obtained in 30% PEG 8000, 0.1 M sodium cacodylate pH 6.5 and 0.2 M sodium acetate (condition No. 28 of Crystal Screen I) after about three months; however, the crystals were not suitable for data collection. While trying to reproduce these crystals, clusters of very thin plates were obtained after 3–4 d. These plates had a clear resemblance to the crystals of the free KH3 domain grown previously (Backe *et al.*, 2004). However, Additive Screens 1 and 2 from Hampton Research with 30% PEG 8000, 0.1 M MES pH 6.5 and 0.05 M sodium acetate gave, in addition to the thin multi-crystals, a new crystal type that again appeared after around three months using almost all the various additives. These crystals were rod-shaped and grew to maximum dimensions of about 0.34 × 0.08 × 0.08 mm (Fig. 1). As described below, these crystals do indeed contain a complex between the KH3 domain and the 15-mer of DNA. Unfortunately, to date we have not been able to reproduce these crystals.

2.2. Data collection and processing

Diffraction data were obtained on the JSBG beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). A single rod-shaped crystal was picked directly from the drop using a thin nylon loop and flash-cooled, without further cryoprotection, in a cold (100 K) nitrogen stream. The crystal-to-detector distance was 230 mm and a total of 60° of data were collected with 0.5° oscillation per frame, starting at an oscillation angle determined using the program *STRATEGY* (Ravelli *et al.*, 1997). Diffraction data were reduced using *MOSFLM* (Leslie, 1992) and *SCALA* (Collaborative Computational Project, Number 4, 1994). The crystals belong to the trigonal space group $P3_121$, with unit-cell parameters $a = 54.0$, $c = 149.9$ Å. Data-collection statistics are summarized in Table 1.

2.3. MicroPIXE measurements

The calculated molecular weights of a single KH3 molecule and the 15-mer ssDNA are 8900 and 4390 Da, respectively. Calcu-

lation of the Matthews coefficient suggests several possibilities for the content of the asymmetric unit. Two KH domains alone would give a V_M of 3.54 Å^3 Da^{-1} and two KH domains in complex with DNA would give V_M = 2.84 Å^3 Da^{-1}, whereas three KH domains without or with DNA would give V_M = 2.36 or 2.02 Å^3 Da^{-1}, respectively. Since initial molecular-replacement trials using the crystal structure of the protein alone (Backe *et al.*, 2004) failed and no heavy-atom derivatives could be obtained despite extensive soaking trials, a micro-analytical technique named microPIXE (proton-induced X-ray emission; Garman, 1999) was used to determine the exact stoichiometry of the KH3–DNA complex.

Two types of crystals were used for the microPIXE technique, one thought to be constituted of the KH3 domain alone and one of a putative KH3–ssDNA complex, both of which grew in the same drop (Fig. 1). The microPIXE technique involves using a 2–3 MeV proton beam of 1 µm in diameter to induce characteristic X-ray emission from crystals dried onto a 2 µm thick Mylar film held in vacuum. These characteristic X-rays are then detected using a solid-state lithium-drifted silicon detector which has high energy resolution. The proton beam is scanned spatially in X and Y; by placing software windows round the X-rays in the spectrum associated with particular elements and sorting these events into an $X–Y$ grid, projection maps can be obtained of all elements heavier than neon that are present in the sample. Of particular interest are the sulfur signal, which reflects the content of methionines or cysteines in the protein, and the phosphorus signal, which reflects the DNA content. Quantitative information is obtained by collecting spectra at three or four points on the crystal and also on the backing foil. These spectra are analysed using the program *GUPIX* (Maxwell *et al.*, 1995) to extract the areal density of each element of interest in the sample. The number of atoms of each element per protein molecule can then be computed from the ratio of the signal from that element compared with that from sulfur, combined with a knowledge of the number of S atoms per molecule, which is available from the protein primary sequence. This internal sulfur calibration allows much more accurate quantitation (normally approximately ±10%, at best ±6%) than would be possible if absolute measurements were necessary.

Since the putative complex crystals were grown in MES buffer, which contains sulfur, it was very important to wash the crystals

Table 1
Crystallographic data and data-collection statistics.

Values in parentheses are for the highest resolution bin.

Beamline/detector	ID14-EH4, ESRF/ADSC Q4R
Wavelength (Å)	0.939
Exposure time per image (s)	5
Space group	$P3_121$
Unit-cell parameters (Å)	$a = 54.00$, $c = 149.95$
Mosaicity (°)	0.3
Oscillation range (°)	0.5
Molecules per asymmetric unit	3 KH3 domains and 1 15-mer DNA molecule
Matthews coefficient (Å^3 Da^{-1})	2.0
Solvent content (%)	38.8
Resolution range (Å)	40.0–2.40 (2.53–2.40)
Total No. reflections	35888
Unique No. reflections	10500
Average redundancy	3.4 (3.3)
Completeness (%)	99.7 (98.3)
R_{merge}† (%)	5.7 (33.6)
I/SD‡	14.1 (3.3)

† $R_{merge} = \sum_{hkl} \sum_i |I(hkl)_i - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i \langle I(hkl)_i \rangle$. ‡ I, intensity; SD, standard deviation.

free of MES prior to mounting them on the Mylar foil. Measurements were made using the Scanning Proton Microbeam Facility in the Department of Materials at the University of Oxford (Grime *et al.*, 1991) on three different crystals: two washed rod-shaped crystals of putative protein in complex with ssDNA (crystals I and II) and a washed plate-shaped putative crystal of uncomplexed protein which had grown in the same crystallization drop but had the same morphology as other crystals of uncomplexed protein grown previously (crystal III) (Backe *et al.*, 2004). All crystals were transferred from their growth drop to separate microbridges containing mother liquor. They were then moved into MilliQ water and agitated, each undergoing six or seven transfers into clean water in approximately 5 min to ensure thorough washing. All three crystals were then placed onto separate 2 µm thick Mylar films. The films had previously been pretensioned over a 1 cm hole in aluminium target holders and then glued to the holders. The microPIXE sulfur maps (Figs. 2*a* and 2*c*) clearly define the shape of the protein crystals.

Analysis of the microPIXE experiments gave a phosphorus to sulfur atomic ratio of 4.4 (±0.3) (three points on the crystal measured at 4.44, 4.35 and 4.43) (Fig. 1) and 4.7 (±0.4) (three points measured at 4.78, 4.45 and 4.83) for the washed putative crystals I and II, respectively; the values are in agreement within errors. Note that statistical errors are significantly smaller than the estimated systematic errors. The washed putative uncomplexed crystal III (see Fig. 2) had a phosphorus to sulfur ratio of 1.1 (±0.4) × 10^{-2}. This result is an

average of two point measurements. At the third point the phosphorus concentration was below the minimum detectable limit. The Mylar backing foil has a small contaminant phosphorus concentration (see Fig. 2d) and the very low phosphorus composition obtained for this crystal most likely arises from this contamination, rather than from the presence of any phosphorus in the crystal.

The results show that crystals I and II contained ssDNA and that crystal III contained very little or none. The measured ratio of phosphorus to sulfur in the washed putative complexed crystals was 4.4 (±0.3) and 4.7 (±0.4), respectively. This agrees very well with the theoretical ratio of 4.67 corresponding to a stochiometry of three protein molecules per 15-mer ssDNA, taking into account the fact that the synthetic oligo has only 14 phosphates.

2.4. Self-rotation function and non-crystallographic symmetry

The results from the microPIXE experiments led us to re-examine the X-ray data. Further evidence for a crystal asymmetric unit containing three KH3 domains per 15-mer of DNA came from self-rotation calculations on the native data performed with *GLRF* (Tong & Rossmann, 1997). Fig. 3 shows a stereographic projection of the $\kappa = 180$ and $40°$ section for the native data, revealing strong evidence for the presence of a pseudo-ninefold non-crystallographic axis ($\kappa = 40°$) parallel to the crystallographic threefold screw axis, suggesting the possibility of a protein trimer in the asymmetric unit.

2.5. Molecular replacement

Molecular-replacement trials were carried out using *MOLREP* (Vagin & Teplyakov, 1997). The search model used was obtained by superimposing the single-stranded DNA in the NMR solution structure of the FUSE-binding protein KH domain (KH4; Braddock *et al.*, 2002) onto our crystal structure of the free KH3 domain (Backe *et al.*, 2004). Using a maximum resolution of 4.0 Å and the program default values for the minimum resolution and search radius, the cross-
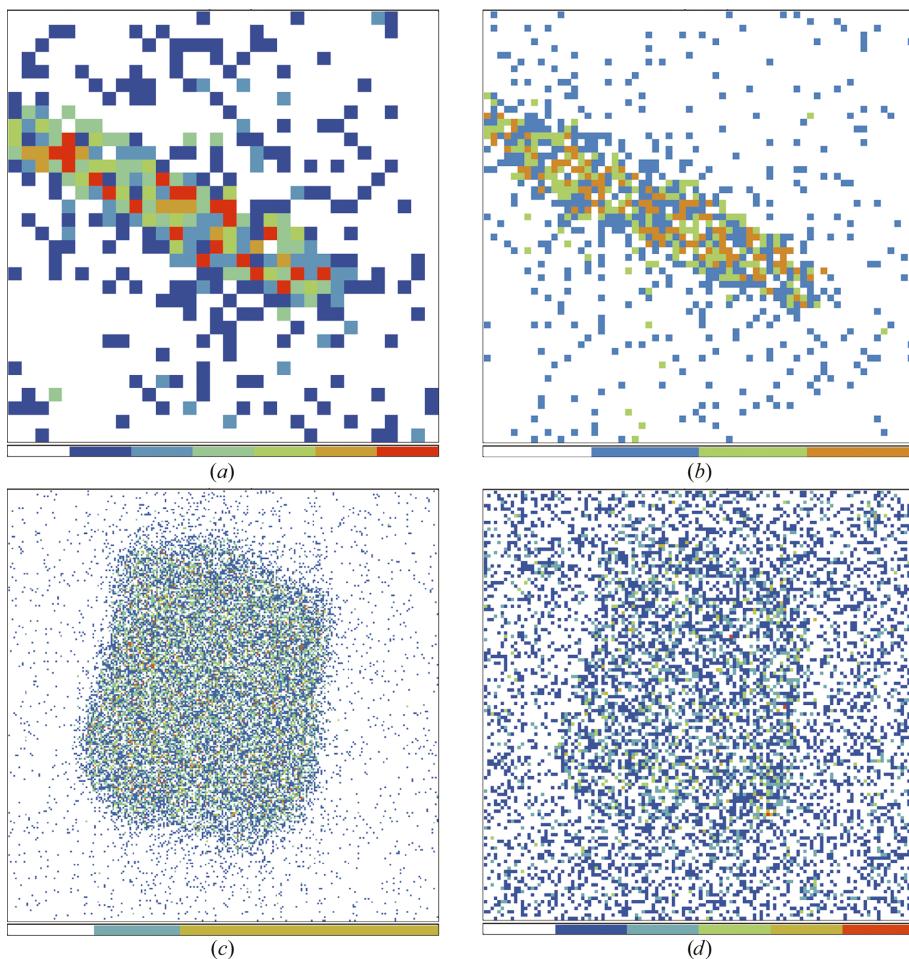

Figure 2
Two-dimensional microPIXE scans giving elemental maps of the crystal. The figure shows the colour-coded intensity of the emitted X-rays for a particular energy (wavelength) corresponding to a unique element, over a grid which encompasses the crystal. The colours range from blue to red, representing low- and high-intensity emission, respectively. Elemental plots (150 × 150 µm scan) are shown for the crystal containing the complex: (a) sulfur, (b) phosphorus. Elemental plots (250 × 250 µm scan) are shown for the crystal containing the free KH3 domain: (c) sulfur, (d) phosphorus. In the sulfur maps, the outline of the crystal is clearly visible in both crystal types. In the case of the phosphorus maps, a significant level of phosphorus can only be detected in the crystal of the complex.

Figure 1
Rod-shaped crystals of the third KH domain of hnRNP K in complex with ssDNA. The thin plates growing in clusters are crystals of the free KH3 domain.
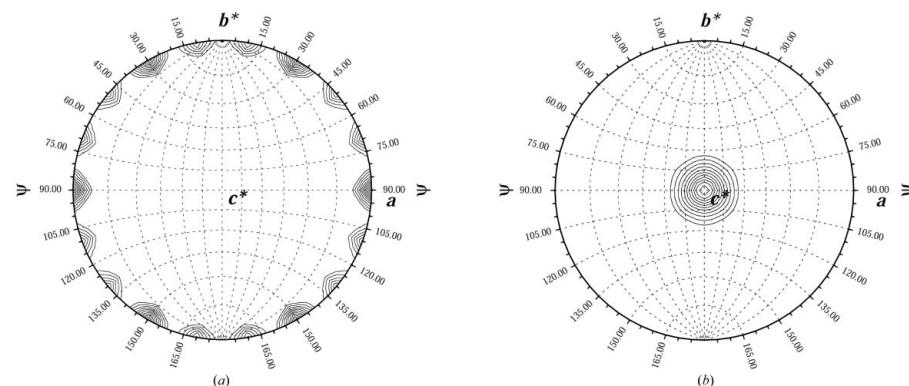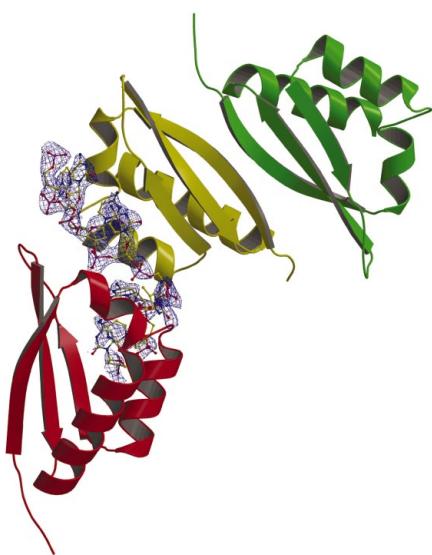

Figure 3
Stereographic projection of (a) the $\kappa = 180$ and (b) the $\kappa = 40°$ sections of the self-rotation function calculated using data in the range 10.0–2.4 Å and an integration radius of 20 Å. The peaks indicate a pseudo-ninefold non-crystallographic axis parallel to the crystallographic threefold screw axis.

Figure 4
Ribbon representation of the three KH3 domains in the asymmetric unit (coloured red, yellow and green) and an omit map showing the electron density around the single-stranded DNA drawn at the 1σ level. Only two of the KH3 domains (red, yellow) interact with the DNA. The third KH3 domain (green) interacts with the second (yellow) *via* the strand $\beta 1$ edges of their respective β-sheets, an interaction also seen in crystals of the uncomplexed protein (Backe *et al.*, 2004). Figure created using *BOBSCRIPT* (Esnouf, 1997) and rendered with *RASTER3D* (Merritt & Murphy, 1994).

rotation function calculation gave no clear peaks. Since the space group was not known, both $P3_121$ or $P3_221$ were used in the subsequent translation search. $P3_121$ gave better results, indicating that this was probably the correct space group. The resulting maps were not very clear and we therefore removed the DNA from the model and performed a rigid-body and a normal refinement in *REFMAC* (Murshudov *et al.*, 1997). This lowered the R_{work} (R_{free}) from 55.5% (53.4%) to 46.1% (49.8%) at 2.4 Å resolution. Although this solution looked promising, we still could not see the DNA in the $F_o - F_c$ difference map. Subsequently, we repeated the molecular replacement using just one copy of the KH3 domain that resulted from the refinement carried out in *REFMAC* as search model. The cross-rotation function calculation gave, among others, three peaks that differed in α angle by approximately 40°, thus being in good agreement with the self-rotation function. These three peaks were then used in the translation function. After fixing the best solution for the first translation function, clear peaks were obtained for a second and a third translation function. The third and last fitted model had a correlation coefficient of 0.386 and an R value of 0.529 (see Table 2).

Table 2
Translation function in the molecular-replacement search.

Values in parentheses correspond to the highest incorrect peaks.

α (°)	β (°)	γ (°)	X_{frac}	Y_{frac}	Z_{frac}	Peak height	CC	R value
0.00	0.00	0.00	0.261	0.318	0.218	14.4 (11.4)	0.223 (0.174)	0.588 (0.607)
38.54	3.95	0.00	0.062	0.323	0.997	28.8 (11.8)	0.303 (0.196)	0.561 (0.601)
77.48	2.83	4.76	0.829	0.170	0.773	57.3 (18.4)	0.386 (0.275)	0.529 (0.575)

A map calculated with this molecular-replacement solution showed clear difference density for some of the DNA.

3. Results and discussion

Fig. 4 shows a ribbon diagram of the three KH domains in the asymmetric unit together with an omit map (Brünger *et al.*, 1998) for the single-stranded DNA contoured at 1σ. The structure reveals that two KH3 domains are bound to successive sequences on the DNA, while the third KH3 domain does not contact the DNA. This result is entirely consistent with the PIXE measurements. Each of the two DNA-bound KH3 domains interact with four bases. The remaining nucleotides are not visible in the electron density, probably because they do not interact with the protein and are consequently flexible. Interestingly, the two KH3 domains bound to DNA are related by a non-crystallographic screw transformation almost parallel to the z axis with a rotation of 162.175° and a z translation of 17.3 Å. This screw axis, with a rotation angle close to a multiple of 1/9 of a full rotation (40°), superimposed on the crystallographic threefold axis accounts for the apparent ninefold symmetry observed in the self-rotation function.

A full description of the structure will be given elsewhere.

References

Baber, J. L., Levens, D., Libutti, D. & Tjandra, N. (2000). *Biochemistry*, **39**, 6022–6032.
Backe, P. H., Ravelli, R. B. G. & Cusack, S. (2004). Submitted.
Bomsztyk, K., Van Seuningen, I., Suzuki, H., Denisenko, O. & Ostrowski, J. (1997). *FEBS Lett.* **403**, 113–115.
Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D. & Clore, G. M. (2002). *Nature (London)*, **415**, 1051–1056.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., 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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002). *Nature Rev. Mol. Cell. Biol.* **3**, 195–205.
Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. (1993). *Annu. Rev. Biochem.* **62**, 289–321.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Garman, E. (1999). *Structure*, **7**, R291–R299.
Grime, G. W., Dawson, M., Marsh, M., McArthur, I. C. & Watt, F. (1991). *Nucl. Instrum. Methods Phys. Res. B*, **54**, 52–63.
Jancarik, J., & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**, 27–33.
Makeyev, A. V. & Liebhaber, S. A. (2002). *RNA*, **8**, 265–278.
Maxwell, J. A., Teesdale, W. J. & Campbell, J. L. (1995). *Nucl. Instrum. Methods B*, **95**, 407.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Michael, W. M., Eder, P. S. & Dreyfuss, G. (1997). *EMBO J.* **16**, 3587–3598.
Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. (1996). *Mol. Cell. Biol.* **16**, 2350–2360.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M. & Hentze, M. W. (1997). *Cell*, **89**, 597–606.
Ostareck-Lederer, A., Ostareck, D. H., Cans, C., Neubauer, G., Bomsztyk, K., Superti-Furga, G. & Hentze, M. W. (2002). *Mol. Cell. Biol.* **22**, 4535–4543.
Ostareck-Lederer, A., Ostareck, D. H. & Hentze, M. W. (1998). *Trends Biochem. Sci.* **23**, 409–411.
Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. (1997). *J. Appl. Cryst.* **30**, 551–554.
Takimoto, M., Tomonaga, T., Matunis, M., Avigan, M., Krutzsch, H., Dreyfuss, G. & Levens, D. (1993). *J. Biol. Chem.* **268**, 18249–18258.
Tomonaga, T. & Levens, D. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 5830–5835.
Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.