Nucleic Acids Research, 1994, Vol. 22, No. 19 3861–3865

Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap binding protein

Naoyuki Kataoka, Mutsuhito Ohno, Kenji Kangawa[1], Yusuke Tokoro and Yoshiro Shimura*
Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606 and [1]Department of Biochemistry, Miyazaki Medical College, Miyazaki 889-16, Japan

Received July 8, 1994; Revised and Accepted August 16, 1994 DDBJ accession no. D32002

ABSTRACT

It has been shown that the monomethylated cap structure plays important roles in nuclear events. The cap structure has been implicated in the enhancement of pre-mRNA splicing. More recently, this structure has also been suggested to facilitate RNA transport from the nucleus to the cytoplasm. We have previously identified and purified an 80kD Nuclear Cap Binding Protein (NCBP) from a HeLa cell nuclear extract, which could possibly mediate these nuclear activities. In this report, we describe cloning of complementary DNA (cDNA) encoding NCBP. The partial protein sequences of NCBP were determined, and the full-length cDNA of NCBP was isolated from HeLa cDNA libraries. This cDNA encoded an open reading frame of 790 amino acids with a calculated molecular mass of 91,734 daltons, which contained most of the determined protein sequences. However, the protein sequence had no significant homology to any known proteins. Transfection experiments demonstrated that the epitope-tagged NCBP, transiently expressed in HeLa cells, was localized exclusively in the nucleoplasm. Similar experiments using a truncated NCBP cDNA indicated that this nuclear localization activity is conferred by the N-terminal 70 amino-acid region.

INTRODUCTION

It has been well accepted that a monomethylated cap structure, $m^7G(5')ppp(5')N$, blocking the 5′ termini of many eukaryotic mRNAs, is important for many cellular functions. In the cytoplasm, the cap structure stimulates translation by facilitating ribosome binding to mRNA (1). A 25 kilodalton cap binding protein (eIF-4E) is involved in this stimulation (reviewed in 1–3).

There has been cumulative evidence indicating that the cap structure also plays important roles in nuclear events. The *in vitro* splicing reaction is inhibited by cap analogs or if uncapped mRNA precursor (pre-mRNA) is used as a substrate (4–8). The formation of the spliceosome was shown to be cap dependent (7). We have previously demonstrated both *in vitro* and in *Xenopus* oocyte nuclei that when pre-mRNAs contain two introns within a single molecule, the enhancing effect of the cap structure seems to be restricted to the splicing reaction of the intron proximal to the structure (6, 8). More recently, the cap structure has been suggested to facilitate export of RNA polymerase II-transcribed RNAs from the nucleus to the cytoplasm (9–12). It is highly likely that nuclear cap binding protein(s) is involved in these nuclear events as is the case with translation.

Several nuclear cap binding proteins from HeLa cells have been described (13–16, 11). However, the biological functions of these proteins remain to be elucidated. We have previously identified and purified an 80 kilodalton cap binding protein from a HeLa nuclear extract, and designated as NCBP (15, 16). This protein is a good candidate for the cap binding protein involved in mRNA splicing and/or RNA export, as already suggested (16, 11). It is anticipated that further characterization of NCBP may lead to better understanding of the biological function of the cap structure. In this report, we describe cloning of complementary DNA (cDNA) encoding NCBP. We also show that NCBP is efficiently localized in the nucleus upon transfection of the cDNA into HeLa cells.

MATERIALS AND METHODS

Protein sequencing

NCBP was purified from HeLa cells as described previously (16). The purified protein (20 μg) was digested with lysil-endopeptidase and trypsin, and the peptides were fractionated by reverse-phase HPLC. Forty-six peptide fractions were isolated and analyzed by the gas-phase sequencer (model 470A, Applied Biosystems).

Reverse transcription and polymerase chain reaction (RT-PCR)

Reverse transcription was carried out using GeneAmp® RNA-PCR kit (Perkin-Elmer Cetus Instruments) and HeLa cell poly(A)$^+$ RNA. Amplification of the cDNA sequence was executed in the PCR thermocycler (Perkin-Elmer Cetus Instruments) by 40 cycles of denaturation (94°C, 1min.), annealing (46°C, 2min.), and extension (72°C, 5min.), followed by the final extension (72°C, 10min.) and soak (4°C, 10 min.). The amplified fragment was digested with *Eco*RI and cloned into pSP73 (Promega). Sequences of the oligonucleotides used for the RT-PCR primers are as follows.

*To whom correspondence should be addressed

primer 1: (5′)-GCGAATTCGAYGTNCCNAAYCCNAA-(3′)
primer 2: (5′)-ATGAATTCTTRAANACYTCRTGRAA-(3′)

N, R and Y represent all bases, purine and pyrimidine, respectively. Primers 1 and 2 correspond to amino-acid sequences of DVPNPN in sense and FHEVFK in antisense orientations, respectively. They also contain an *Eco*RI site close to the 5′ end.

Isolation of NCBP cDNA

The λgt10 HeLa cell cDNA libraries were screened by plaque hybridization using the digoxigenine-labeled NCBP cDNA fragment. DNA labeling, hybridization and detection were achieved as recommended by the manufacturer (Boehringer Mannheim).

Plasmid constructions

The oligonucleotides coding for the 12 amino acid sequence from the T7 gene10 protein (T7 epitope) were synthesized, annealed and inserted at the N-terminus of the NCBP coding region in the cDNA (at the *Bam*HI site). The resulted cDNA fragment was cloned into *Xba*I site of pEF-BOS (17). Sequences of the oligonucleotides corresponding to T7 epitope are as follows;

T7-sense:
(5′)-GATCCATGGCTAGCATGACTGGTGGACAGCAAATGGGTCG-(3′)
T7-anti:
(5′)-GATCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATG-(3′)

For the experiment of the truncated NCBP, the *Bam*HI-*Bgl*II fragment of the cDNA was used.

Northern and Southern blot analyses

For Northern blot analysis, HeLa cell poly(A)$^+$ RNA (2μg per lane) was electrophoresed in a 1% formamide-agarose gel and blotted onto Hybond N nylon membrane (Amersham). Hybridization was performed in 50% formamide, 5×SSC, 1% blocking reagent (Boehringer Mannheim), 0.1% N-laurylsalkosyl and 0.02% SDS at 42°C for 12 hours using ^{32}P-labeled NCBP cDNA as a probe. After hybridization, the membrane was rinsed twice in 2×SSC/0.1% SDS at room temperature, and washed twice in 0.2×SSC/0.1% SDS at 42°C for 15min., followed by autoradiography. For Southern blot, HeLa cell genomic DNA (2 μg per lane) was digested with restriction enzymes, and electrophoresed in a 0.7% agarose gel. The following procedures are the same as Northern blot.

Cell culture and transfection

HeLa cells were cultured at 37°C in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. The DNA-calcium phosphate coprecipitation method (18) was used for the transfection. In a typical assay, HeLa cells cultured in a 60 mm dish were transfected with 10μg of the test plasmid. After 24 hours, the cells were washed three times with phosphate-buffered saline (PBS, pH7.4) and subjected to immuno-staining or Western blot analysis.

Immuno-staining of HeLa cells

Immuno-staining of HeLa cells was performed as described (19). The fixed HeLa cells were first incubated with either anti hnRNP C monoclonal antibody (4F4) or anti T7 tag monoclonal antibody diluted 1:1000, followed by incubation with fluorescein isothiocyanate-conjugated F(ab′)$_2$ fragment of goat anti mouse immunoglobulins (DAKO A/S) diluted 1:300. Localization of the antigens was visualized by the fluorescence microscope (model AXIOPHOTO FL; KARL ZEISS).

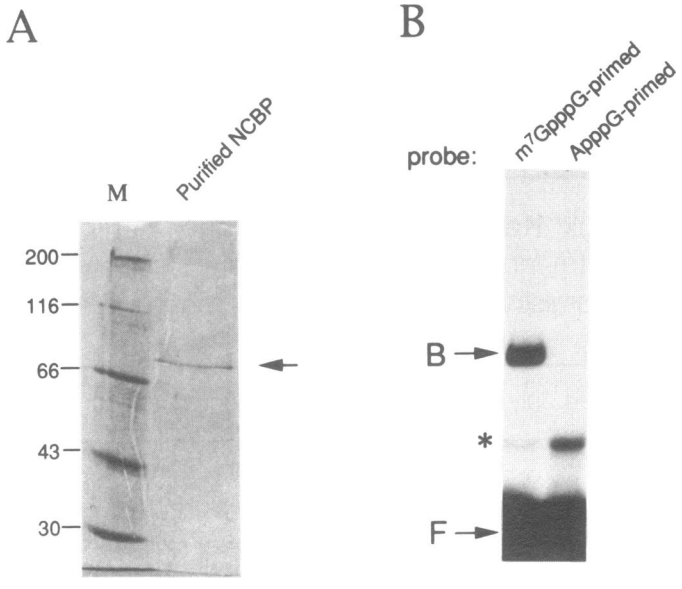

Figure 1. Purified NCBP and its cap binding activity. (**A**) The purified NCBP preparation (150ng) was electrophoresed in a 10% SDS-polyacrylamide gel and visualized by coomassie brilliant blue staining. The molecular size markers are shown on the left side of the figure. The band of 80 kD NCBP is indicated by an arrow. (**B**) Cap binding activity of purified NCBP (10ng) was assayed by the gel mobility shift assay using either m^{7}GpppG- or ApppG-primed RNA probe as described previously (15, 16). The RNA bands not shifted and shifted by NCBP are indicated by F and B, respectively. The band indicated by an asterisk corresponds to the free probe of an alternative secondary structure.

RESULTS AND DISCUSSION

Cloning of a cDNA encoding NCBP

As described previously, we identified an 80 kilodalton cap binding protein (NCBP) in a HeLa cell nuclear extract by the gel mobility shift assay using a short capped RNA as a probe, and established the biochemical procedure to purify the protein to near homogeneity (15, 16).

To investigate the property and the function of NCBP, an attempt was made to clone a cDNA encoding NCBP. For this purpose, approximately 20 μg of NCBP was purified from 1×10^{11} HeLa cells. The purified fraction was essentially homogeneous, as judged by SDS/PAGE followed by coomassie staining, thereby a single band of approximately 80 kilodaltons was visible (Fig. 1A). Moreover, the purified NCBP still retained specific binding activity to the cap structure. As shown in Fig. 1B, the purified protein could bind efficiently to the monomethylated cap structure (m^{7}GpppG-primed probe) but not at all to a pseudo-cap structure in which the blocking nucleotide is adenosine instead of guanosine (ApppG-primed probe) as judged by the gel shift assay.

The purified NCBP was digested with lysil-endopeptidase and trypsin, and the peptides were fractionated by reverse-phase HPLC. The sequences of many peptides were determined (see Fig. 2). A number of degenerate PCR primers were synthesized according to the peptide sequences, and used in pair-wise combinations to amplify the NCBP cDNA fragments by reverse

```
MSRRRHSDEN DGGQPHKRRK TSDANETEDH LESLICKVGE KSACSLESNL   50
EGLAGVLEAD LPNYKSKILR LLCTVARLLP EKLTIYTTLV GLLNARNYNF  100
GGEFVEAMIR QLKESLKANN YNEAVYLVRF LSDLVNCHVI AAPSMVAMFE  150
NFVSVTQEED VPQVRRDWYV YAFLSSLPWV GKELYEKKDA EMDRIFANTE  200
SYLKRRQKTH VPMLQVWTAD KPHPQEEYLD CLWAQIQKLK KDRWQERHIL  250
RPYLAFDSIL CEALQHNLPP FTPPPHTEDS VYPMPRVIFR MFDYTDDPEG  300
PVMPGSHSVE RFVIEENLHC IIKSHWKERK TCAAQLVSYP GKNKIPLNYH  350
IVEVIFAELF QLPAPPHIDV MYTTLLIELC KLQPGSLPQV LAQATEMLYM  400
RLDTMNTTCV DRFINWFSHH LSNFQFRWSW EDWSDCLSQD PESPKPKFVR  450
EVLEKCMRLS YHQRILDIVP PTFSALCPAN PTCIYKYGDE SSNSLPGHSV  500
ALCLAVAFKS KATNDEIFSI LKDVPNPNQD DDDDEGFSFN PLKIEVFVQT  550
LLHLAAKSFS HSFSALAKFH EVFKTLAESD EGKLHVLRVM FEVWRNHPQM  600
IAVLVDKMIR TQIVDCAAVA NWIFSSELSR DFTRLFVWEI LHSTIRKMNK  650
HVLKIQKELE EAKEKLARQH KRRSDDDDRS SDRKDGVLEE QIERLQEKVE  700
SAQSEQKNLF LVIFQRFIMI LTEHLVRCET DGTSVLTPWY KNCIERLQQI  750
FLQHHQIIQQ YMVTLENLLF TAELDPHILA VFQQFCALQA             790
```

Figure 2. Deduced amino-acid sequence of the full length NCBP cDNA. Amino-acid sequence is shown in single-letter code. The peptide sequences obtained from the purified protein are shaded.

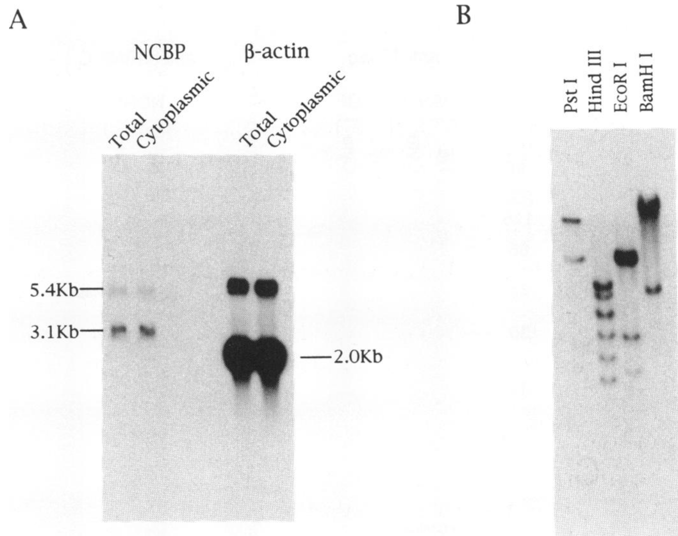

Figure 3. Northern and Southern blot analyses. (**A**) Northern blot analysis of NCBP transcript. Both total and cytoplasmic poly(A)$^+$ RNAs from HeLa cells were analyzed using the NCBP and β-actin cDNA probes. (**B**) Southern blot analysis of the genomic DNA from HeLa cells using the NCBP cDNA probe.

transcription PCR. One combination of primers yielded an amplified DNA fragment of approximately 160 base-pairs. Cloning and sequencing of the fragment revealed that it represented a part of the NCBP coding sequence, since the sequence from another peptide was able to be identified in the fragment.

Using this cDNA fragment as a probe, the HeLa cell cDNA libraries were screened, and a 2.9 kb cDNA was cloned. This cDNA could encode an open reading frame of 790 amino acids with a calculated molecular weight of 91,734 daltons, which contained most of the identified peptide sequences. The deduced protein sequence is shown in Fig. 2. This cDNA is highly likely to encode the whole coding region of NCBP, since an in-frame stop codon is found upstream of the assigned initiating methionine, and also the size of the cDNA (2.9 kb) is consistent with the size of the major NCBP transcript (3.1 kb; see Fig. 3A for Northern analysis), assuming that the poly A tail is about 200 nucleotides long. There are two peptide sequences that could not be assigned in the coding sequence. These peptides were likely to originate from minor contaminating protein(s) in the purified NCBP preparation. It remains to be clarified whether the minor contaminants represent protein(s) associated with NCBP. The N terminus of the protein is highly charged, and within the C terminal one third of the protein, two short stretches of aspartic acid are found. However, no significant homology was found with any known proteins in the sequence data bases (GenBank, EMBL and SWISS-PROT).

NCBP mRNA is expressed mainly as a 3.1kb transcript in HeLa cells (Fig. 3A), which is consistent with the size of the full-length cDNA as already mentioned above. A minor 5.4kb transcript, also seen in the figure, is reproducibly detected in both total and cytoplasmic poly(A)$^+$ RNAs from HeLa cells. Further characterization of this longer transcript suggested that it contains 2.3kb extra sequence in the 3′ terminal region as compared with the shorter transcript (data not shown). Genomic Southern blot analysis indicated that the copy number of the NCBP gene is less than a few copies per HeLa cell genome (Fig. 3B).

Localization of NCBP in HeLa cells

Since NCBP is a candidate for the cap binding protein involved in mRNA splicing and/or RNA export, it is of interest to elucidate it's localization in the cell. Although subcellular fractionation already suggested that this protein is localized mainly in the nucleus (15, 16), it is important to confirm the biochemical result which is sometimes ambiguous.

For this purpose, we carried out immuno-staining of HeLa cells, which enabled us to localize the protein in the cells more precisely than biochemical fractionation. Since specific antibody against NCBP has not yet been obtained, we took advantage of the epitope-tag. The full length NCBP cDNA was placed downstream of the elongation factor EF-1α promoter (17), and oligonucleotides corresponding to the 12 amino-acid sequence from the T7 phage gene10 were inserted at the N-terminus of the NCBP coding region. This plasmid was transfected into HeLa cells by the calcium phosphate method. After 24 hours, the nuclear and cytoplasmic extracts were prepared from the transfected HeLa cells, and analyzed by SDS/PAGE and Western blot using anti-T7 tag monoclonal antibody in order to monitor the expression of the epitope-tagged NCBP in the HeLa cells. Separately, the transfected HeLa cells were fixed, permeabilized, and immuno-stained with the same antibody.

As shown in Fig. 4A, the tagged NCBP was detected in both extracts. NCBP was detected more abundantly in the cytoplasmic extract than in the nuclear extract. However, much of NCBP detected in the cytoplasmic extract is likely to have leaked out from the nucleus during the extract preparation, since we also detected in the cytoplasmic extract much of hnRNP C protein, which is known to reside exclusively in the nucleus (20). The molecular mass of the epitope-tagged NCBP is approximately 80 kilodaltons in the SDS gel as is the case with native NCBP, which confirms that the cloned cDNA contains the whole coding region of NCBP in spite of the fact that the molecular weight

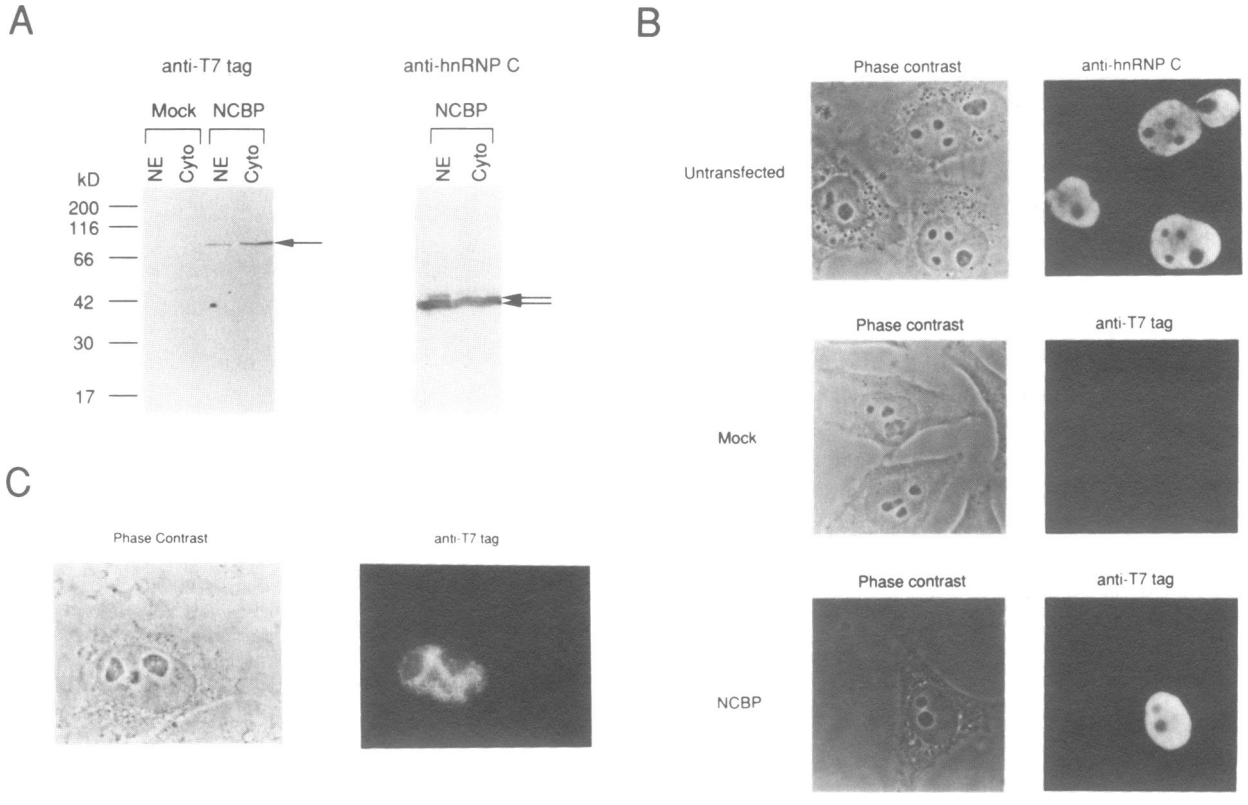

Figure 4. Localization of the epitope-tagged NCBP. (**A**) Western blot analysis of the epitope-tagged NCBP. HeLa cells were transfected by either the plasmid expressing the epitope-tagged NCBP or the vector itself. After 24 hours, the cells were harvested and fractionated into the nuclear and cytoplasmic fractions as described previously (22). The fractions were subjected to 15% SDS-PAGE and analyzed by Western blot using either anti T7 tag antibody or anti hnRNP C monoclonal antibody, 4F4 (20). (**B**) Immuno-staining of the transfected HeLa cells. HeLa cells transfected as in (A) were fixed and stained by either anti-hnRNP C monoclonal antibody (Upper panel) or anti-T7 tag monoclonal antibody (Middle and Lower panels). (**C**) Nuclear localization of the truncated NCBP. A truncated NCBP cDNA containing only the N terminal 70 amino acids was used for the same experiment as in (B).

calculated from the cDNA sequence is slightly larger (92 kilodaltons).

As shown in Fig. 4B, if the transfected HeLa cells were immuno-stained with the epitope antibody, nucleoplasmic staining (nuclear staining with the exception of the nucleoli) but not cytoplasmic staining was observed in many cells (bottom row), whereas the cells transfected with the vector itself yielded no detectable signal (middle row). This nucleoplasmic staining pattern was very similar to the case of hnRNP C protein (top row). These results strongly suggest that NCBP is exclusively localized in the nucleoplasm, although we cannot exclude the possibility that the pattern observed here may not precisely reflect the localization of endogenous NCBP due to the overexpression of the protein.

If a truncated NCBP cDNA, which contained only the N terminal 70 amino-acid region instead of the whole coding region, was used in the same experiment, the expressed protein was still efficiently localized in the nucleus (Fig. 4C). This indicates that the nuclear localization signal resides within the first 70 amino-acid region of NCBP. It has been proposed that two short stretches of basic amino acids separated by 10 amino-acid spacer constitute a nuclear localization signal (21). In fact, such sequence is found near the N terminus of NCBP (amino-acid residue 3 to 20; RRRHSDENDGGQPHKRRK in single-letter amino-acid code).

It was suggested by Izaurralde *et al.* that NCBP (or CBP80 according to their nomenclature) might be involved in RNA export from the nucleus (11). However, no direct evidence has ever been presented. The biological function of NCBP remains to be clarified. We are currently preparing specific antibodies against the recombinant NCBP. It is anticipated that these antibodies will help clarifying the function of NCBP in the near future.

ACKNOWLEDGEMENTS

We thank Drs G.Dreyfuss for the monoclonal antibody 4F4, H.Handa for HeLa cDNA libraries, and T.Kunisada for the plasmid pEF-BOS. We also thank the members of our laboratory for helpful advice and comments on the manuscript. This work was supported by grants from the Ministry of Education and Science, Japan, and from the Inamori Foundation.

REFERENCES

1. Shatkin,A.J. (1985) Cell, 40, 223–224.
2. Rhoads,R.E. (1988) Trends Biochem. Sci., 13, 52–56.
3. Sonenberg,N. (1988) Prog. Nucleic Acid Res. Mol. Biol., 35, 174–207.
4. Konarska,M.M., Padgett,R.A. and Sharp,P.A. (1984) Cell, 38, 731–736.
5. Edery,I. and Sonenberg,N. (1985) Proc. Natl. Acad. Sci. USA, 82, 7590–7594.

6. Ohno,M., Sakamoto,H. and Shimura,Y. (1987) Proc. Natl. Acad. Sci. USA, 84, 5187–5191.
7. Patzelt,E., Thalmann,E., Hartmuth,K., Blaas,D. and Kuechler,E. (1987) Nucleic Acids Res., 15, 1387–1399.
8. Inoue,K., Ohno,M., Sakamoto,H. and Shimura,Y. (1989) Genes Dev., 3, 1472–1479.
9. Hamm,J. and Mattaj,I.W. (1990) Cell, 63, 109–118.
10. Dargemont,C. and Kühn,L.C. (1992) J. Cell Biol., 118, 1–9.
11. Izaurralde,E., Stepinski,J., Darzynkiewicz,E. and Mattaj,I.W. (1992) J. Cell Biol., 118, 1287–1295.
12. Jarmolowski,A., Boelens,W.C., Izaurralde,E. and Mattaj,I.W. (1994) J. Cell Biol., 124, 627–635.
13. Patzelt,E., Blaas,D. and Kuechler,E. (1983) Nucleic Acids Res., 11, 5821–5835.
14. Rozen,F. and Sonenberg,N. (1987) Nucleic Acids Res., 15, 6489–6500.
15. Ohno,M. and Shimura,Y. (1990) Methods Enzymol. 181, 209–215
16. Ohno,M., Kataoka,N. and Shimura,Y. (1990) Nucleic Acids Res., 18, 6989–6995.
17. Mizushima,S. and Nagata,S. (1990) Nucleic Acids Res., 18, 5322.
18. Wigler,M., Sweet,R., Sim,G.K., Wold,B., Pellicer,A., Lacy,E., Maniatis,T., Silverstein,S. and Axel,R. (1979) Cell, 16, 777–785
19. Harlow,L. and Lane,D. (1988) Antibodies:A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
20. Choi,D.Y. and Dreyfuss,G. (1984) J. Cell Biol., 99, 1997–2004
21. Robbins,J. ,Dilworth,S.M., Laskey,R.A. and Dingwall,C. (1991) Cell, 64, 615–623
22. Lee,A.W.K. and Green,R.M. (1990) Methods Enzymol.181, 20–30