GRSF-1: a poly(A)⁺ mRNA binding protein which interacts with a conserved G-rich element

Zuwei Qian and Jeffrey Wilusz*

Department of Microbiology and Molecular Genetics, UMDNJ – New Jersey Medical School and Graduate School of Biomedical Sciences, South Orange Avenue, Newark, NJ 07103, USA

Received March 7, 1994; Revised and Accepted May 16, 1994

GenBank accession no. U07231

ABSTRACT

Computer predictions identified similarities to a 14-base G-rich element in numerous mRNAs at a variety of locations. A Northwestern screening strategy was used to obtain a cDNA clone from a HeLa cell library using the G-rich RNA element as a probe. A cellular protein (called GRSF-1), which was encoded by this cDNA, binds RNAs containing the G-rich element. GRSF-1 was distinct from DSEF-1, a nuclear protein we have previously identified that interacts with the G-rich element, based on differences in molecular weight and partial peptide maps, as well as the lack of cross-reactivity with GRSF-1 specific monoclonal antibodies. Using indirect immunofluorescence microscopy, we localized GRSF-1 to the cytoplasm. *In vivo* UV cross-linking further demonstrated that GRSF-1 was bound to poly(A)⁺ mRNA in living human cells. Western blot analysis revealed four cytoplasmic proteins which expressed GRSF-1 specific epitopes. GRSF-1 contains three potential RNA recognition motifs and two auxiliary domains. Curiously, the domain organization of GRSF-1 is similar to the RNA binding proteins PUB1, ELAV, HuD, Hel-N1, mcs94-1 and RBP9.

INTRODUCTION

In addition to control at the level of synthesis, the expression of many mRNAs is influenced at several post-transcriptional levels including polyadenylation (1), splicing (2, 3), stability (4), transport and localization (5). Over the past 10 years or so, considerable progress has been made in defining the *cis*-acting elements and *trans*-acting factors involved in some aspects of post-transcriptional regulation. An emerging theme from these studies is the role of RNA–protein interactions in regulated post-transcriptional events. The identification and characterization of RNA binding proteins, therefore, has become an important step towards an understanding of many of these events.

Nuclear post-transcriptional events are well documented to be influenced by protein–RNA interactions. Splicing may be regulated by a group of specialized factors (SR proteins) which are capable of committing pre-mRNA to splicing pathways (6). In addition, the SR proteins can interact with other splicing factors

(7) and, in conjunction with hnRNP A proteins (8), may promote splice site selection. Processing of the 3' end of pre-mRNAs is also mediated by RNA–protein interactions. The interplay between two *cis*-acting elements (AAUAAA and a U-rich downstream element) and *trans*-acting factors (CPSF and CstF) may determine the location and efficiency of 3' end cleavage and polyadenylation (9–11).

Sequence-directed mRNA–protein interactions may also influence RNA stability. The 3' untranslated region (UTR) of many short lived mRNAs contains an AUUUA element which specifically interacts with proteins and destabilizes mRNAs (12–19). The half-life of the transferrin receptor mRNA is also influenced by a protein–RNA interaction within the 3' UTR (20–22). Sequences within the mRNA coding region also regulate mRNA half-life by sequence-specific protein–RNA interactions (23–25).

Translation is also influenced by RNA binding proteins. An interesting paradigm is the conserved 28 base RNA sequence in the 5' UTR of ferritin mRNA which binds a 98 kDa protein whose role in modulating translation is dependent on specific physiological conditions (22, 26, 27). Similarly, a 52 kDa host protein interacts with the 5' UTR of polio virus mRNA and may influence viral protein synthesis (28). Sequence-specific RNA–protein interactions involving the 3' UTR and poly(A) tail also influence the translation efficiency of mRNA (29–33).

RNA–protein interactions also affect mRNA transport and localization. In *Xenopus*, the interaction between a 340-nucleotide sequence in the 3' UTR of vg1 mRNA and a 69 kDa protein directs the vegetal localization of the transcript (34, 35). Cooperation between *staufer* (a double-stranded RNA binding protein) and the 3' UTR of bicoid mRNA is required for anterior and posterior pole localization of the mRNA in *Drosophila* oocytes (36, 37). Finally, in human immunodeficiency virus, the binding of Rev protein to the Rev response element is critical for cytoplasmic transport of partially spliced viral mRNA (38).

All of the examples cited above illustrate the significance of RNA–protein interactions in the control of gene expression. A range of RNA binding proteins exists to provide more generalized as well as specific functions. At one end of this range are the general RNA binding proteins, best illustrated by several of the hnRNP proteins (39). Nuclear hnRNP proteins bind RNA with

*To whom correspondence should be addressed

limited specificity and may play more global roles in RNA metabolism. At the other end of the range lie highly specific RNA binding factors which recognize primary sequence and/or secondary structure. These proteins are likely to have specialized roles in RNA processing. The identification of sequence-specific RNA-protein interactions, therefore, may lead to further insights into mechanisms of post-transcriptional regulation.

In this study, a computer analysis was used to identify similarities to a 14 base G-rich element on numerous mRNAs. This G-rich element may influence the fate of this subset of mRNAs through sequence-specific RNA-protein interactions. We used a Northwestern screening strategy to identify and characterize a cDNA clone encoding a cytoplasmic mRNA binding protein called GRSF-1 (G-Rich Sequence Factor-1). GRSF-1 contains three RNA recognition motifs and has a structural organization very similar to a set of previously characterized RNA binding proteins. GRSF-1 protein was capable of binding to RNAs *in vitro* in a G-rich element-dependent manner and interacted with poly (A)⁺ mRNA in living human cells.

MATERIALS AND METHODS

cDNA cloning and determination of DNA and amino acid sequences

cDNA library screening. Detailed methodology used in the identification of cDNAs encoding GRSF-1 is outlined elsewhere (40). In order to obtain full length clones, a HeLa S-3 cell cDNA expression library cloned in the Uni-ZAP XR vector (Stratagene) was screened using a ZQ50-1.0 RNA probe (40) following the manufacturer's recommendation with minor modifications. The resulting cDNA clone contained a 2.7 kb *EcoRI*-*XhoI* fragment representing an apparently full length cDNA fragment based on sequence information (see below).

cDNA sequence and amino acid sequence analysis. Nucleotide sequence of the cDNA insert was determined by the dideoxynucleotide chain termination method (41) using the DNA Sequenase Kit ver. 2.0 (United States Biochemicals) from both DNA strands. T3 primer (ATTAACCTCACTAAAG), M13-20 primer (GTAAACGACGGCCAGT) and a set of internal primers were used for sequence analysis. Amino acid sequence predictions were carried out using EUGENE software on a UNIX machine. Amino acid sequence homology searches were performed at the National Center for Biotechnology Information using the basic local alignment search tool (BLAST) network service (42) within a non-redundant data base comprising the Brookhaven Protein Data Bank (June 1993 release), SWISS-PROT 27.0, PIR 38.0 and CDS translations from GenBank (R) release 79.0. Amino acid sequence alignments were performed using the MacDNASIS Pro ver. 2.0 software (Hitachi Software Engineering Co.).

RNA, DNA and protein analyses

RNA probes. GEM50 and GEM4 RNAs (59 and 65 bases, respectively) were transcribed *in vitro* and purified by successive NH₄Ac/ethanol precipitations as previously described (40). RNA probes used for cDNA library screening were transcribed by T7 RNA polymerase from a *SmaI* linearized plasmid pZQ50-1.0 which was constructed by inserting the 1 kb ZQ50 cDNA insert into pGEM3 at the *EcoRI* site (40).

Genomic Southern blot analysis. This was carried out essentially as described (43). Approximately 5 µg of genomic DNA were digested with *EcoRI* and fractionated on a 1% agarose gel. After transfer to a NYTRAN PLUS membrane (Schleicher & Schuell), the blot was hybridized with a DNA probe which was prepared from pZQ50-1.0 (40) using the random primers DNA labeling system (BRL). The hybridized blot was washed using a final stringency of 0.1×SSC/1% SDS at 63°C, and bands were visualized by autoradiography.

Northern blot analysis. This was performed as described (43) with some modifications. Briefly, 2–3 µg of poly(A)⁺ RNA or 20–30 µg of total RNA from HeLa cells were treated with formaldehyde and formamide before fractionation on a 1% formaldehyde agarose gel. After transfer to a NYTRAN membrane (Schleicher & Schuell), RNA was hybridized with the same DNA probes used for genomic Southern blot analysis in a solution containing 50% formamide at 45°C overnight. The blot was washed in 0.1×SSC, 0.1% SDS at 50°C and bands were visualized by autoradiography.

Western blot analysis. This was performed as described (43) with some modifications. Proteins were fractionated by SDS-PAGE and transferred to a nitrocellulose membrane (Schleicher & Schuell) in 25 mM Tris, 198 mM glycine, 20% methanol and 0.1% SDS (44) at 10 V overnight. The membrane was blocked in blocking buffer (100 mM Tris-Cl, pH7.5, 0.9% NaCl supplemented with 5% nonfat dry milk, 0.05% NP40, 0.02% sodium azide) at room temperature for 1 hr and incubated with 1 ml hybridoma supernatant diluted in 30 ml blocking buffer. Following extensive washings with Tris-buffered saline, the membrane was incubated with anti-mouse IgG alkaline phosphatase conjugate (Sigma Chemical Company), washed again and developed with Lumi-Phos 530 (Boehringer Mannheim).

RNA-protein crosslinking. *In vitro* UV cross-linking was performed as previously described (45). *In vivo* UV cross-linking was done according to Dreyfuss *et al.* (46) as described below. HT1080 human fibroblast cells were grown to near confluence in DME medium supplemented with 10% fetal bovine serum. Cells were washed several times in PBS and were irradiated with a 15 W germicidal UV light in the culture dish. The irradiated cells were swollen in ice-cold 10 mM Tris-Cl (pH 7.4), 10 mM NaCl, 1.5 mM MgCl₂ containing 0.001% aprotinin and 0.0001% leupeptin (Boehringer Mannheim) before homogenization by four passages through a 25-gauge needle. The cytoplasmic fraction was collected by low-speed centrifugation and adjusted to 1 mM EDTA, 1% mercaptoethanol, and 0.5% SDS. After heating at 95°C for 5 min and quenching on ice, LiCl was added to 0.5 M and the cytoplasmic extract was incubated for 15 min with oligo(dT)-cellulose (Boehringer Mannheim) with agitation on a rocking platform. The mixture was then packed in a column and washed extensively with binding buffer. The poly(A)⁺ material was eluted in 10 mM Tris-Cl (pH 7.4), 1 mM EDTA and 0.05% SDS and precipitated with 3 volumes of ethanol overnight. The pelleted poly(A)⁺ material was digested with RNase A before analysing by SDS-PAGE. The protein bands were detected by Western blot analysis (see above) using anti-GRSF-1 monoclonal antibody.

Partial proteolysis by V8 protease. GEM50 RNA (approximately 3×10^7 cpm) was UV cross-linked to either 4 μ g of purified recombinant GRSF-1 protein or 14.5 μ l (~ 100 μ g) HeLa nuclear extract (45) and digested with RNase A before loaded onto an 8% acrylamide–SDS gel. Cross-linked DSEF-1 and GRSF-1 specific protein bands were excised following autoradiography. The gel slices were further cut into several equal size pieces and loaded onto a 15% acrylamide–SDS gel. Each gel slice was overlaid with 38 μ l 2 $\times$ gel loading buffer containing increasing concentrations of *Staphylococcus aureus* V8 protease (Boehringer Mannheim). Protein samples were allowed to run to the middle of the stacking gel and electrophoresis was stopped for 30 min to allow in-gel digestion (47).

Expression of recombinant GRSF-1 from *E. coli*

To facilitate subcloning of the GRSF-1 cDNA fragment, the polylinker region of the expression vector pGEX2T was expanded by replacing the region between *Bam*HI and *Eco*RI restriction sites with a new polylinker constructed using the oligomer GATCCCTCGGGTCGACGGTACCTCGAG and its appropriate complement, which resulted in new plasmid pGEX2TZQ. A 2.7 kb *Ava*I cDNA fragment was directionally cloned into pGEX2TZQ at the *Ava*I site (resulting plasmid pGEX2TZQ-2.7). This 2.7 kb fragment contained almost the entire GRSF-1 open reading frame and the 3' UTR with only the 5' UTR and sequences corresponding to the first 7 amino acids deleted, resulting in a net loss of 5 amino acids at the N-terminus. *E. coli* strain HB101 was transformed with pGEX2TZQ-2.7 and expression of the GST fusion protein was induced by IPTG. GST fusion proteins were purified by passing the cell sonicate through a glutathione–sepharose affinity column (Pharmacia Biotech Inc.) (48). The purified GST fusion proteins were then digested with thrombin (which cleaved at the junction of the two fusion

partners) and the GST portion was removed by passing the reaction mixture through a glutathione–sepharose column and collecting the flow-through fraction.

Indirect immunofluorescence microscopy

Human fibroblast cells HT1080 were grown in DME medium supplemented with 10% fetal bovine serum on a chamber slide (Lab-Tek Products). Cells were fixed using three different fixation methods including incubation in 3% formaldehyde in PBS at room temperature for 20 min, 1% glutaraldehyde at room temperature for 1 hr, or in methanol/acetone [50%/50%, (v/v)] at room temperature for 2 min. Slides were then transferred to pre-chilled acetone for 5 min, rinsed thoroughly in PBS, and blocked in PBS containing 3% BSA at 37°C for 1 hr. Hybridoma supernatant containing the anti-GRSF-1 monoclonal antibody was diluted (1:20) in 3% BSA and applied to the slide for 12 hrs at 4°C. The slide was washed four times in PBS containing 0.1% Tween-20 for 2 min each before anti-mouse IgG/F(ab')₂-FITC (Boehringer Mannheim), diluted 1:25 in 3% BSA, was added. Following incubation at room temperature for 3 hr, the slide was washed once in 0.1% Tween-20 followed by three washes in PBS and photographed under an Olympus fluorescence microscope using Kodak Tri-X pan 400 black and white film.

Monoclonal antibody production

Antigen was purified as a β -galactosidase–GRSF-1 fusion protein from a lysogenic strain we have established previously (40). Total proteins were separated by SDS-PAGE, detected with copper chloride (49) and electroeluted using the ELUTRAP electro-separation system (Schleicher & Schuell). Approximately 500 μ g of β -gal fusion protein were injected into BALB/c mice for immunization. Hybridoma production (49) and screening (50) were conducted as described.

Table 1. A sample representation of DSEF-1 binding site similarities in mRNAs

mRNA	Sequence ¹	Match ²	Location ³
SV40 late transcription unit	G G G G G A G G U G U G G G ⁴	14/14	DSE ⁵
Human c-fgr proto-oncogene	G G G G G A G G U G G G G c	11/14	5' UTR
Human B-myb (myb homologue)	G G G G G A u G U c U c G G	11/14	5' UTR
Human c-abl oncogene	G G c G G c G G U G a G G G	11/14	5' UTR
Human cAMP-responsive element binding protein (CREB)	G G c G G A G G U G U a G u	11/14	5' UTR
Human bcl-2 proto-oncogene	G c G G G A G a U G U G G G	12/14	Coding Region
Human B-raf oncogene	G G a G G A G G U G U G G a	12/14	Coding Region
Human alpha-interferon	G c a G G A G G U G U G G G	12/14	Coding Region
Human apoE	G a G G A c G U G U G c G	11/14	Coding Region
Human fmr-1	G G U G G A a G U G c G G G	11/14	Coding Region
Human c-erbA-1 (thyroid hormone receptor) ⁶	G G G G G A G c U G U G u c	11/14	Coding Region
Human c-abl oncogene	c G G G G A G G U G U a c G	11/14	Coding Region
Human colony stimulating factor (CSF-4)	c G a G G A G G U G U c G G	11/14	Coding Region
Human c-jun proto-oncogene	G G G G G A G G g u U G G G	12/14	3' UTR
Human beta-interferon	a G G G G A G c U G U G G G	12/14	3' UTR
Human beta-nerve growth factor	G a G G G A G G U G U G a u	11/14	3' UTR
Human c-fms proto-oncogene	u G G G a A G G U G a G G G	11/14	3' UTR
Human ets-2 oncogene	G G G G G c G G U a g G G G	11/14	3' UTR
Human activating transcription factor (ATF-a)	G G G G G A a u U G U G u G	11/14	3' UTR

¹The 14 base DSEF-1 binding site similarities. Lower case letters indicate variations from the prototypic SV40 sequence element.

²Number of bases matched to the DSEF-1 binding site.

³Location of the sequence element on the mRNA.

⁴The sequence that was used to search the GenBank database.

⁵Functionally important element located downstream of the SV40 late poly(A) site.

⁶One of five similar sequence elements in c-erbA-1 mRNA.

Nucleotide sequence accession number

The DNA and amino acid sequences for human GRSF-1 have been submitted to the GenBank data base and have been assigned the accession number U07231.

RESULTS

G-rich binding site similarities are present on many mRNAs

We have previously identified the RNA binding site of the cellular 50 kDa DSEF-1 protein to consist of a 14 base G-rich element, 5' GGGGGAGGUGUGG 3' (45). This G-rich element was also able to mediate efficient processing of the SV40 late polyadenylation signal (45). Using the 14 base G-rich DSEF-1 binding site as probe, a computer-aided search for similar sequence elements in the GenBank database was performed. By allowing a maximum of three mismatches, DSEF-1 binding site similarities were found not only in several polyadenylation signals at positions both upstream and downstream of the cleavage site (51), but also at a variety of locations on numerous mRNAs and pre-mRNAs. These similarities were found at least seven times more frequently than expected on a random basis, with over 7000 entries in the primate data base alone. A partial list of G-rich elements is presented in Table 1. The G-rich elements were found in mRNAs encoding onco-proteins, transcription factors and other biologically important proteins. ApoE and FMR-1 mRNAs, for example, encode proteins implicated in Alzheimer's disease (52, 53) and fragile X syndrome (54, 55) respectively. It may also be noteworthy that these G-rich elements were also found in intronic regions. These observations raised an interesting possibility that the G-rich element may be a general *cis*-acting element involved in several aspects of RNA post-transcriptional regulation.

Molecular characterization of a cDNA clone encoding a G-rich element specific RNA binding protein

Since the DSEF-1 protein binds the 14 base G-rich sequence and may be involved in mediating efficient polyadenylation, the wide-

spread occurrence of G-rich elements on a variety of mRNAs suggests that other proteins which interact with this sequence may also be present in the cell. We have recently developed a Northwestern strategy for screening cDNA libraries. Using this approach, we have isolated a 1 kb partial cDNA clone which encodes a peptide which interacts with the G-rich element (40). In order to obtain a full length cDNA clone, an RNA probe corresponding to the partial cDNA clone was used to screen a cDNA library derived from the HeLa S-3 cell line by direct nucleic acid hybridization. One clone, which contained an apparently full length 2.7 kb cDNA insert, was isolated and subjected to DNA sequencing from both strands. Computer analysis indicated that the cDNA encoded an open reading frame (ORF) which contained a translation start codon whose sequence context (TCCATGG) matched very well with the consensus (A/GCCATGG) (56, 57). The putative open reading frame encodes a 424 amino acid polypeptide of 48 kDa which has an isoelectric point of 5.43 and is rather hydrophilic in nature (overall hydrophobicity -0.41). The cDNA clone (GenBank accession number U07231) has a long 3' UTR which occupied approximately half of its length (nucleotides 1273-2629) followed by a poly(A) tail beginning at position 2630. The 3' UTR contains two polyadenylation signals (AATAAA) located at positions 1479 and 2604. We have isolated an additional cDNA clone which used the 5' proximal polyadenylation signal (data not shown), suggesting both polyadenylation signals are functional. This additional clone also contained a different 5' portion upstream of position 100 (data not shown), indicating that the primary transcripts of this gene may be subjected to alternative RNA splicing.

The GRSF-1 ORF has several interesting domains (Fig. 1A). First, the protein contains three RNA recognition motifs (RRM), designated as RRM1 (residues 95-178), RRM2 (residues 195-275) and RRM3 (residues 346-424). Each RRM contains two highly conserved sequence motifs, RNP1 and RNP2, as shown underlined in Fig. 1B. RRM1 and RRM2 share 70%

A.

B.

CONSENSUS	IYIKGL LFVGN	E L F F D Y	GK RGFAFVXF R K Y I	A I G I V V
RRM1 (95-178)	L <u>IRAQGL</u> PWSCTMEDVLNFFSDCRIRNGENIHFLNRD--GKR <u>RGDALIEM</u> ESEQDVQKALEKHRMYM-QQRYVEVYEINNEDVDA			
RRM2 (195-275)	V <u>VRLRGL</u> PYSCNEKDIVDFAGLNI----VDITFVMDYRGRRK- <u>TGEAYVOE</u> EPEMANQALLKHREEI-GNRYIEIFPSRRNEVRT			
RRM3 (346-424)	F <u>VHMRGL</u> PFQANAQDIINFAPLKP----VRIT--MEYSSSGKA <u>TGEADVHE</u> ETHEDAVAAMLKDRSHV-HHRYIELFLNSCPKGGK-			
	RNP2		RNP1	

Figure 1. GRSF-1 domain organization and RRM sequence alignment. (A) Diagrammatic representation of the RRM and auxiliary domains. Gray boxes indicate RRM, the dark box indicates the helical acidic domain and gradient box indicates the alanine-rich domain. Numbers indicate the amino acid positions range of each domain. (B) GRSF-1 RRM domain alignment. RNP-1 and RNP-2 consensus sequences and conserved residues (58,85) are indicated on top of the alignment and are underlined in each respective RRM. Consensus sequences and residues are indicated on top of the alignment.

similarity and 37% identity, RRM1 and RRM3 share 59% similarity and 28% identity, and RRM2 and RRM3 share 67% similarity and 43% identity. This is consistent with the general observation that RRMs within a single protein typically show low identity (58). Second, an acidic domain (residues 296–337) between the RRM2 and RRM3 was predicted with glutamic acid (31%) and aspartic acid (2.4%) representing 33.4% of the amino acid content. A Chou–Fasman protein structure prediction suggested extensive amphipathic α -helical structure within this acidic domain, reminiscent of the activation domain in some of the transcription factors (59–61). Finally, towards the N-terminus of GRSF-1 there is a stretch (residues 27–55) rich in alanine (38%).

To further characterize and determine the identity of the cloned protein, the cDNA clone was inserted into a bacterial protein expression vector in-frame with glutathione S-transferase. Recombinant fusion protein was expressed and purified using a glutathione affinity column (Fig.2A). The 79 kDa fusion protein represented approximately 80% of the purified protein fraction. This purified GST–GRSF-1 fusion protein was then subjected to thrombin proteolysis which specifically cleaved at the GST–recombinant protein junction (Fig.2B) and the GST peptide was removed by collecting the flow through of a glutathione affinity column (Fig.2B). The purified recombinant protein migrated with apparent molecular weight of 53 kDa in SDS-PAGE. The disparity with the calculated molecular weight (48 kDa) is likely due to the overall negative charge of the protein.

Northern blot analysis using either total RNA or polyadenylated RNA from HeLa cells indicated two predominant mRNAs at ~1.7 kb and ~2.8 kb (Fig.3A). This was consistent with the observation that multiple polyadenylation signals were apparently

used to generate different cDNA clones (see above). The size of the 2.8 kb mRNA could represent the mRNA for the cDNA clone specifying the full length GRSF-1 which used the distal polyadenylation signal.

To test the phylogenetic conservation of the GRSF-1 gene, genomic DNA from different sources was digested with *EcoRI* and used in a Southern blot analyzed with probes corresponding to the coding region of the GRSF-1 cDNA clone (Fig.3B). The blot was washed under relatively stringent conditions, $0.1 \times \text{SSC}/1\% \text{ SDS}$ at 63°C , to ensure only highly homologous sequences would be detected. DNA fragments which hybridized with the GRSF-1 gene were found in hamster, calf, mouse, rat as well as human genomes, whereas no hybridizing fragments were observed in *Saccharomyces cerevisiae* or *Drosophila melanogaster* genomic DNAs. Our results suggested that GRSF-1 is a relatively conserved gene in vertebrates.

Using purified β -gal–GRSF-1 fusion proteins from a lysogenic strain (40) as source of antigen, two monoclonal antibodies (1G2 and 6H11) were developed which specifically interacted with GRSF-1 recombinant proteins or *in vitro* translation products (data not shown). When used in western analysis of total cytoplasmic HeLa cell protein, both monoclonal antibodies detected a series of four polypeptides of 56 kDa, 60 kDa, 70 kDa and 74 kDa (Fig.3C). We assume that the 56 kDa protein represents the GRSF-1 protein due to its apparent molecular weight. These data suggest that GRSF-1 is a member of a family of at least four cytoplasmic proteins.

RNA binding properties of GRSF-1

As one might expect, the purified GST–GRSF-1 protein would interact with RNA in a G-rich element-dependent manner because

Figure 2. Purification of recombinant GRSF-1 proteins from *E. coli*. (A) Purification of GST–GRSF-1 fusion protein. The total cell sonicate (lane Total Protein) was prepared from IPTG-induced *E. coli* containing pGEX2TZQ-2.7 and passed through a glutathione affinity column (lane Flow-Thru), the column was washed extensively with PBS buffer (lane Wash) and GST fusion proteins were eluted with glutathione elution buffer (lane Elution). Size markers are indicated on the right side of the gel. The major band around 79 kDa (indicated by the arrow) is the GST–GRSF-1 fusion protein. (B) The purified GST–GRSF-1 fusion protein was cleaved with thrombin which separated the 29 kDa GST portion and the 53 kDa recombinant GRSF-1 portion (indicated by the arrow) (lane BEFORE COLUMN). The digested protein mixture was then passed through a glutathione affinity column and the 53 kDa GRSF-1 portion was collected in the flow through fraction (lane AFTER COLUMN). M, size standards.

Figure 3. Molecular characterization of GRSF-1 specific DNA, mRNA and protein products. (A) Northern blot analysis. Total RNA (left lane) or poly(A)⁺ RNA (right lane) from HeLa cells was fractionated on a 1% denaturing agarose gel, transferred onto a NYTRAN membrane and probed with radiolabeled ZQ50 DNA. Relative migration of size standards is shown on the right. (B) Southern blot. 5 μ g of *Eco*RI digested total genomic DNA from different organisms (indicated above the lanes) was electrophoresed, blotted and hybridized with the ZQ50 probe. Relative migration of size standards is shown on the right side of the panel. (C) Western blot analysis. Proteins from a HeLa cell cytoplasmic S100 fraction were electrophoresed on an 8% SDS-PAGE gel, transferred to a nitrocellulose membrane and detected by the anti-GRSF-1 monoclonal antibody. Relative molecular weights of the four proteins are indicated on the right.

the cDNA clone encoding GRSF-1 protein was isolated by Northwestern screening using RNA probes containing the G-rich element. Indeed, as shown in Fig. 4, a 59 base RNA containing the G-rich element (GEM50) specifically bound the GST-GRSF-1 fusion protein and was nearly quantitatively co-precipitated by the GRSF-1 specific monoclonal antibody 6H11 (panel B, lane GEM50). A negative control RNA without the G-rich element (GEM4) was not co-precipitated by the antibody (lane GEM4), demonstrating a sequence-specific interaction between the GEM50 RNA element and the GRSF-1 recombinant protein. Additional data using variants of the G-rich element indicated that the interaction with GRSF-1 was somewhat tolerant to point mutations within the G-rich element. All of the G-rich element point mutations tested retained $\geq 30\%$ binding (data not shown). These results were very similar to those obtained for the interaction of DSEF-1 with RNAs (data not shown).

GRSF-1 is distinct from the DSEF-1 protein

Recombinant GRSF-1 protein is approximately the same size as DSEF-1 protein and has very similar RNA binding properties. The three observations outlined below using GRSF-1 and DSEF-1 proteins cross-linked to radioactively labeled GEM50 RNA by UV light, however, indicated they were indeed independent proteins. First, the two proteins migrated slightly differently on a high resolution gel (Fig. 5B). Second, cross-linked proteins were subjected to immunoprecipitation by the GRSF-1 specific monoclonal antibody 1G2. As shown in Fig. 5A, 1G2 failed to precipitate the cross-linked DSEF-1 protein (lane 3), but cross-linked GRSF-1 protein was efficiently precipitated (lane 4).

A. RNA CONSTRUCTS

B. CO-IMMUNOPRECIPITATION

Figure 4. Recombinant GST-GRSF-1 protein binds short RNAs containing a G-rich element. (A) Diagrammatic representation of the RNAs used in co-immunoprecipitation experiment. Numbers indicate the size of the RNAs. (B) Co-immunoprecipitation experiment. Equal counts of GEM4 or GEM50 RNA were incubated with purified GST-GRSF-1 fusion protein to allow complex formation. The GRSF-1 specific monoclonal antibody was added and immune complexes were precipitated by protein A-positive *Staphylococcus aureus* cells. Proteins were removed by phenol extraction and RNAs were recovered by ethanol precipitation and analyzed on 5% denaturing polyacrylamide gel. Arrow indicates GEM50 RNA.

Figure 5. GRSF-1 and DSEF-1 are distinct proteins. (A) Either DSEF-1 in HeLa nuclear extract (NE, lane 1) or recombinant GRSF-1 (RP, lane 2) was cross-linked to radioactively labeled GEM50 RNA and treated with RNase A before loaded onto SDS-PAGE. In lanes 3–6, the indicated cross-linked proteins were immunoprecipitated using anti-GRSF-1 antibody (lanes 3 and 4) or a control antibody (lanes 5 and 6) prior to electrophoresis. Recombinant GRSF-1 (upper arrow) and DSEF-1 (lower arrow) are indicated. (B) UV cross-linking of either DSEF-1 protein (left lane) or GRSF-1 protein (right lane) to radiolabeled GEM50 RNA. Cross-linked proteins were analyzed on a high resolution 6% SDS acrylamide gel. Rightward arrow indicates DSEF-1 protein. Leftward arrow indicates recombinant GRSF-1 protein.

Identical results were observed with monoclonal antibody 6H11 (data not shown).

Finally, peptide mapping of the two cross-linked proteins was performed using V8 protease from *Staphylococcus aureus*. The two proteins were digested with increasing amounts of V8 protease to ensure partial proteolysis, and peptide patterns were analyzed by SDS-PAGE (Fig. 6). Significant differences can be seen in the patterns of partial proteolysis products obtained for cross-linked DSEF-1 versus GRSF-1. We conclude, therefore, that DSEF-1 and GRSF-1 are different proteins that share a common RNA binding specificity.

GRSF-1 is a cytoplasmic poly(A)⁺ mRNA binding protein

The determination of the subcellular localization of GRSF-1 protein may provide a clue for its role in RNA metabolism. Monoclonal antibody 1G2 was used to localize GRSF-1 in human fibroblasts by indirect immunofluorescence microscopy. As shown in Fig. 7, GRSF-1 was exclusively detected in cytoplasm with a speckled staining pattern. A similar staining pattern was obtained by three different methods of fixation prior to immunolocalization, suggesting that the cytoplasmic staining pattern was not an artifact of fixation. In a parallel experiment, a monoclonal antibody specific for hnRNP A1 protein (62) exclusively stained the nucleus and exhibited a diffuse pattern throughout the nucleoplasm (data not shown). Western analysis of HeLa cell cytoplasmic proteins probed with monoclonal antibody 1G2 identified four protein species ranging from 56 to 74 kDa (Fig. 3C). This raised an interesting possibility that there might be five different proteins (GRSF-1, DSEF-1, 60 kDa, 70 kDa and 74 kDa) which belong to the same protein family capable of binding the G-rich element.

Since G-rich sequence elements are on many mature mRNAs (Table 1), we decided to test the possibility that GRSF-1 was

Figure 6. Partial proteolysis analysis of cross-linked GRSF-1 and DSEF-1 proteins. DSEF-1 and GRSF-1 proteins were cross-linked to radioactive GEM50 RNA, gel purified, digested with different amount of V8 protease and run on a 15% SDS-PAGE. The gradient wedges on top of the gel denote increasing V8 protease concentration: lane 1 and 7, 0 µg/ml; lane 2 and 8, 0.26 µg/ml; lane 3 and 9, 2.6 µg/ml; lane 4 and 10, 26 µg/ml; lane 5 and 11, 130 µg/ml; lane 6 and 12, 260 µg/ml.

a cytoplasmic mRNA binding protein. Human fibroblast cells grown to near confluence on petri dishes were irradiated with UV light which cross-linked closely associated proteins to RNA

Figure 7. Indirect immunofluorescence microscopy indicated GRSF-1 proteins were exclusively localized in the cytoplasm. Human fibroblast cells were fixed, incubated with anti-GRSF-1 monoclonal antibody 1G2 followed by incubation with anti-mouse IgG, F(ab')₂-FITC from goat as a secondary antibody. The scale bar represents approximately 40 μ m.

Figure 8. GRSF-1 protein is a cytoplasmic poly(A)⁺ mRNA-binding protein. Human fibroblast cells were irradiated with UV light, cytoplasmic RNA was extracted and poly(A)⁺ RNAs with covalently cross-linked proteins were isolated by oligo(dT) chromatography and digested with RNase A. The cross-linked proteins were electrophoresed on an 8% SDS-PAGE gel and analyzed by Western blotting using an anti-GRSF-1 antibody (lane 3). Additional lanes on the blot include a parallel control experiment in which cells were not irradiated with UV light prior to poly(A)⁺ RNA isolation and protein analysis was performed (lane 4), and cytosolic proteins (lanes 1 and 2).

in vivo (46). Poly(A)⁺ cytoplasmic RNAs were then isolated by oligo(dT) chromatography and the eluted poly(A)⁺ fraction, which contained proteins cross-linked to mRNAs, was treated with RNase A and analyzed by SDS-PAGE. The gel was blotted to a nitrocellulose membrane and probed with monoclonal antibody 1G2. This monoclonal antibody was able to detect GRSF-1 and related proteins cross-linked to poly(A)⁺ mRNAs

A.

GRSF-1	94	FLTRAOGLPWSCTMEDVLNFFSDCRIRNGENGIHLLNRDGRGDALIE	143
mcs94-1	10	FVYKLRGLPWSCSVEDVQNFLSDCTIHDGAAGVFIYTTREGROSGEAFVE	59
GRSF-1	144	MESEQDVQKALEKHRMYMGQRYVEVEIN-NE-DVDALMK-SL-QV-KSS	193
mcs94-1	60	LGSEDDVKMALKKDRESMGHRYIEV--FKSHRTMDWVLKHSGPNSADSA	109
GRSF-1	194	PVVNDGVVRLRGLPYSCNEKDIDVDFAGLNIV-D-ITFVMDYRGRRTGE	243
mcs94-1	110	---NDGFYRLRGLPFGCTKEEIVQFFSGLEIVPNGITLPVDPGKI-TGE	159
GRSF-1	244	AYVQFE-EPEMANQALLKHREEIGNRYIEIFPSRRNEVRT--H--VG--S	293
mcs94-1	160	AFVQFASQ-ELAELKLGKHKERIGHRYIEVFKSSQEEVRSYSDPPLKFMS	209
GRSF-1	294	YK--G--KKIASFPTAKYITEPEMVFEHEVNEVPQPMT-AFE-S--EK-	343
mcs94-1	210	VQRPGPYDRPGT--ARRYIG--IV-KQAGLERM-RPGAYSTGYGGYEY	259
GRSF-1	344	---E---I---ELP-KEVP-----EKLPE-A-ADFG--TTSSLHFVHM	393
mcs94-1	260	SGLSDGYGFTTDLFGRDLSYCLSGMYDH--RYGDSEFTVQSTTG-HCVHM	309
GRSF-1	394	RGLPFQANAQDII-NFFAPLKPVRITMEYSSSGKATGEADVHFETHEDAV	443
mcs94-1	310	RGLPYKA-TENDIYNFFSPLNPVRVHIEIGPGRVTEADVFEATHEEAV	359
GRSF-1	444	AAMLKDRSHVHHRYIELFLNSCPGK.....	497
mcs94-1	360	AAMSKDRANMQHRYIELFLNSTTGAS.....	413

B.

Figure 9. The domain organization of GRSF-1 protein is very similar to a class of RRM-type proteins. (A) Amino acid sequence homology between GRSF-1 and mcs94-1 proteins. Single underlines indicate RNP-2 hexamer in each individual RRM. Double underlines designate RNP-1 octamer consensus in each RRM. Similar or identical amino acids between two proteins are linked by vertical lines. (B) Diagrammatic representation of a class of RNA binding proteins which share similar domain organization to GRSF-1 protein.

(Fig.8, lane 3). The 70 and 74 kDa GRSF-1 related proteins were also detected bound to poly(A)⁺ mRNA. No bands were detected when cells were not irradiated with UV light prior to analysis (Fig.8, lane 4), indicating the specificity of the assay. The slight fuzziness and shift of the protein bands in lane 3 compared to the GRSF-1 specific proteins detected in the total cytoplasmic protein fraction in lanes 1 and 2 was due to the heterogeneity of the short RNA molecules which were still attached to the protein after cross-linking and RNase A digestion (71). Based on our indirect immunofluorescence staining and *in vivo* UV cross-linking data, we conclude, therefore, that GRSF-1 protein is a cytoplasmic mRNA binding protein.

DISCUSSION

In this study we report the isolation and analysis of a full length cDNA clone which encodes a protein, GRSF-1, which binds to a common 14 base G-rich element. Molecular characterization indicated that the 424 amino acid GRSF-1 protein contained three potential RNA recognition motifs and two other potentially

interesting domains. Indirect immunofluorescence microscopy and *in vivo* UV cross-linking experiments suggested that GRSF-1 protein was a cytosolic poly(A)⁺ mRNA binding protein.

The amino acid sequence of GRSF-1 protein revealed several interesting features. First, the protein contains three conserved RNA recognition motifs which have been identified as RNA binding domains. RRM domains have approximately 21 conserved amino acid residues spread along an 80–90 amino acid region, with the most conserved regions being the RNP-1 octamer sequence and RNP-2 hexamer sequence (58, 63, 64). Comparison of the three RRM domains of GRSF-1 indicated that they shared low sequence homology amongst themselves. The fact that GRSF-1 contained multiple non-identical RNA binding domains is consistent with the notion that the protein may be able to bind to multiple segments of RNA if each RRM has an independent specificity (58). Each of the four domains of *Xenopus* and yeast PABP, for example, has its own RNA-binding activity and specificity (65, 66). Such proteins can tether different RNA segments and bring into close proximity sequence elements that are otherwise far away from each other. GRSF-1, therefore, may function in RNA packaging, stabilization of RNA secondary structure, or other macromolecular interactions.

Auxiliary domains of RNA-binding proteins may play a role in protein–protein interactions. The GRSF-1 protein contains a highly acidic, α -helical domain located between RRM2 and RRM3. Acidic, α -helical regions of proteins have been shown to be transcriptional activation domains (59–61). This acidic domain, therefore, may allow GRSF-1 to interact with other proteins. A second auxiliary domain is an alanine-rich region (residues 27–55) at the N-terminus. This domain is homologous to an alanine-rich region next to the C-terminus of the *Dictyostelium* ribosomal large subunit P2 protein (67), suggesting a possible role in translation. The exact function and significance of this alanine rich domain of GRSF-1, however, needs to be determined.

Auxiliary domains are also suggested to provide binding specificity (58). The monoclonal antibodies raised against recombinant GRSF-1 have identified three additional proteins (60 kDa, 70 kDa, and 74 kDa respectively). Since the monoclonal antibody was raised against the region which comprises the three RRM domains of GRSF-1, all three proteins may be able to interact with the G-rich element. The size difference of these proteins may suggest divergent auxiliary domains present on each protein which may confer different sequence specificity or protein–protein interactions for members of the GRSF-1 family.

We have introduced a variety of single point mutations in the G-rich element by site-directed mutagenesis. Both DSEF-1 and GRSF-1 proteins exhibited striking similarity in binding to variant G-rich elements, suggesting the RNA sequence specificities of the two proteins are very similar. RRM domains or structural units responsible for such binding specificity, therefore, may be conserved in DSEF-1 and GRSF-1. None of the single base changes completely abolished GRSF-1 protein–RNA interactions (data not shown). GRSF-1, therefore, probably interacts to some degree with the numerous mRNAs which contain G-rich elements detected in our computer search. GRSF-1 protein–RNA interactions may, therefore, occur on many cytoplasmic RNAs, and could potentially play an important role in mRNA packaging, transport, stability and/or translation. The cytoplasmic speckled staining pattern seen with GRSF-1 specific antibodies in indirect immunofluorescence microscopy may indicate a structural organization or compartmentalization of GRSF-1 proteins.

It is of particular interest to note that computer search using the BLAST program against the PIR data bank revealed that GRSF-1 shared a very high degree of similarity (70% over a 350 amino acid region) at the amino acid sequence level with the mcs94-1 protein (Fig.9A). The mcs94-1 protein, which is somewhat similar to mouse nucleolin, was originally identified as a gene located in the *MEN2A* (Multiple Endocrine Neoplasia type 2A) region of human chromosome 10q11.2 (68). While mcs94-1 was originally found to have two RRM domains, our computer analysis indicated the potential existence of a third RRM at the C-terminus. Comparison of the individual corresponding RRM domains between GRSF-1 and mcs94-1 revealed striking similarities. RRM1 of the two proteins shared 80% similarity and 51% identity; RRM2 domains, 82% similarity and 55% identity; while RRM3 sharing 84% similarity and 63% identity. The corresponding regions between RRM2 and RRM3 did not seem to have significant sequence homology. Two groups have recently independently published schemes on the evolutionary classification of RRM domains and RRM-type protein domain organization which reflect functional similarity of these proteins (69, 70). In light of these schemes, it is tempting to speculate that GRSF-1 and mcs94-1 may have similar functions. It is also noteworthy that GRSF-1 can also be classified along with a group of interesting RNA binding proteins such as PUB1 (yeast) (71), ELAV (*Drosophila*) (72), HuD (human) (73) and Hel-N1 (human) (74) and RBP9 (*Drosophila*) (75) which all have three RRM domains and similar domain organization (Fig.9B). The similar domain organization suggests that these proteins may form a family, evolving from a common ancestor through a similar evolutionary pathway.

Phosphorylation of RNA binding proteins has been shown to be critical for their activities in some regulated processes (76–83). Amino acid sequence analysis of GRSF-1 indicated a number of phosphorylation sites. GRSF-1 protein contained three casein kinase II sites (residues 83–88, 203–208 and 317–322), one proline-dependent protein kinase site (residues 187–190), one cGMP-dependent protein kinase site (residues 267–270), one S6 kinase II site (residues 58–63) and two myosin heavy chain kinase sites (residues 284–288 and 403–407) (84). It may be interesting to determine whether GRSF-1 protein is phosphorylated by protein kinases and the biological implications of this phosphorylation.

In this report, we presented the characterization of a cDNA clone encoding an RNA binding protein which was isolated using a Northwestern screening technique. In many instances when studying RNA–protein interactions, the first information obtained is the RNA binding site. By using a Northwestern strategy, one may quickly begin the molecular characterization of these factors which interact with RNA sequence elements. By identifying and characterizing a group of mRNA binding proteins which recognize conserved RNA sequence elements, one may be able to gain insights into some of the important aspects of mRNA processing events.

ACKNOWLEDGEMENTS

We would like to thank Paramjeet S. Bagga for help on indirect immunofluorescence microscopy and sharing of some of his computer analysis data. We thank Fan Chen for assistance on tissue culture techniques and helpful discussions, Marty Marlow for monoclonal antibody development, Daniel Kenan and Jack Keene for assistance with RRM analysis, Yu Fang for helpful

suggestions and Satnam Banga for excellent photography skills. This work was supported by an NIH grant (AI31165) and a grant from the Council for Tobacco Research, USA Inc. to J.W. J.W. is a Pew Scholar in the Biomedical Sciences.

REFERENCES

- Wahle, E. and Keller, W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.
- Mattaj, I.W. (1989) *Cell*, **57**, 1–3.
- Maniatis, T. (1991) *Science*, **251**, 33–34.
- Atwater, J.A., Wisdom, R. and Verma, I.M. (1990) *Annu. Rev. Genet.*, **24**, 519–541.
- Standart, N. (1993) *Molecular Biology Reports*, **18**, 135–142.
- Fu, X.-D. (1993) *Nature*, **365**, 82–85.
- Wu, J.Y. and Maniatis, T. (1993) *Cell*, **75**, 1061–1070.
- Mayeda, A. and Krainer, A.R. (1992) *Cell*, **68**, 365–375.
- Wilusz, J., Shenk, T., Takagaki, Y. and Manley, J.L. (1990) *Mol. Cell. Biol.*, **10**, 1244–1248.
- Weiss, E.A., Gilmartin, G.M. and Nevins, J.R. (1991) *EMBO J.*, **10**, 215–219.
- MacDonald, C., Wilusz, J. and Shenk, T. Submitted for publication.
- Brewer, G. (1991) *Mol. Cell. Biol.*, **11**, 2460–2466.
- You, Y., Chen, C.-Y.A. and Shyu, A.-B. (1992) *Mol. Cell. Biol.*, **12**, 2931–2940.
- Savant-Bhonsale, S. and Cleveland, D.W. (1992) *Genes Dev.*, **6**, 1927–1939.
- Malter, J.S. (1989) *Science*, **246**, 664–666.
- Malter, J.S. and Hong, Y. (1991) *J. Biol. Chem.*, **266**, 3167–3171.
- Gillis, P. and Malter, J.S. (1991) *J. Biol. Chem.*, **266**, 3172–3177.
- Bohjanen, P.R., Petryniak, B., June, C.H., Thompson, C.B. and Lindsten, T. (1991) *Mol. Cell. Biol.*, **11**, 3288–3295.
- Vakalopoulou, E., Schaack, J. and Shenk, T. (1991) *Mol. Cell. Biol.*, **11**, 3355–3364.
- Hentz, M.W., Rouault, T.A., Harford, J.B. and Klausner, R.D. (1989) *Science*, **244**, 357–359.
- Casey, J.L., Hentz, M.W., Koeller, D.M., Caughman, S.W., Rouault, T.A., Klausner, R.D. and Harford, J.B. (1988) *Science*, **240**, 924–928.
- Rouault, T.A., Stout, C.D., Kaptain, S., Harford, J.B. and Klausner, R.D. (1991) *Cell*, **64**, 881–883.
- Chen, C.-Y.A., You, Y. and Shyu, A.-B. (1992) *Mol. Cell. Biol.*, **12**, 5748–5757.
- Bernstein, P.L., Herrick, D.J., Prokipcak, R.D. and Ross, J. (1992) *Genes Dev.*, **6**, 642–654.
- Parker, R. and Jacobson, A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2780–2784.
- Klausner, R.D., Rouault, T.A. and Harford, J.B. (1993) *Cell*, **72**, 19–28.
- Haile, D.J., Hentz, M.W., Rouault, T.A., Harford, J.B. and Klausner, R.D. (1989) *Mol. Cell. Biol.*, **9**, 5055–5061.
- Meerovitch, K., Pelletier, J. and Sonenberg, N. (1989) *Genes Dev.*, **3**, 1026–1034.
- Wharton, R.P. (1992) *Seminars in Dev. Biol.*, **3**, 391–398.
- Evans, T.C., Goodwin, E.B. and Kimble, J. (1992) *Seminars in Dev. Biol.*, **3**, 381–390.
- Richter, J.D. (1991) *BioEssays*, **13**, 179–183.
- Wickens, M. (1992) *Seminars in Dev. Biol.*, **3**, 399–412.
- Sachs, A.B. and Deardorff, J.A. (1992) *Cell*, **70**, 961–973.
- Schwartz, S.P., Aisenthal, L., Elisha, Z., Oberman, F. and Yisraeli, J.K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11895–11899.
- Mowry, K.L. and Melton, D.A. (1992) *Science*, **255**, 991–994.
- St Johnston, D., Beuchle, D. and Nusslein-Volhard, C. (1991) *Cell*, **66**, 51–63.
- St Johnston, D., Brown, N.H., Gall, J.G. and Jantsch, M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10979–10983.
- Cullen, B.R. and Malim, M.H. (1991) *Trends Biochem. Sci.*, **16**, 346–350.
- Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. and Burd, C.G. (1993) *Ann. Rev. Biochem.*, **62**, 289–321.
- Qian, Z. and Wilusz, J. (1993) *Anal. Biochem.*, **212**, 547–554.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.
- Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.
- Qian, Z. and Wilusz, J. (1991) *Mol. Cell. Biol.*, **11**, 5312–5320.
- Dreyfuss, G., Adam, S.A. and Choi, Y.D. (1984) *Mol. Cell. Biol.*, **4**, 415–423.
- Cleveland, D.W., Fischer, S.G., Kirschner, M.W. and Laemmli, U.K. (1977) *J. Biol. Chem.*, **252**, 1102–1106.
- Smith, D.B. and Johnson, K.S. (1988) *Gene*, **67**, 31–40.
- Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.
- Hawkes, R., Niday, E. and Gordon, J. (1982) *Anal. Biochem.*, **119**, 142–147.
- Bagga, P.S., Walton, W. and Wilusz, J. Submitted for publication.
- Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. and Pericak-Vance, M.A. (1993) *Science*, **261**, 921–923.
- Strittmatter, W.J., Weisgraber, K.H., Huang, D.Y., Dong, L.M., Salvesen, G.S., Pericak-Vance, M.A., Schmechel, D.E., Saunders, A.M., Goldgaber, D. and Roses, A.D. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8098–8102.
- Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., Eussen, B.E., van Ommen, G.-J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A. and Warren, S.T. (1991) *Cell*, **65**, 905–914.
- Siomi, H., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1993) *Cell*, **74**, 291–298.
- Kozak, M. (1986) *Cell*, **44**, 283–292.
- Kozak, M. (1989) *J. Cell. Biol.*, **108**, 229–241.
- Bandziulis, R.J., Swanson, M.S. and Dreyfuss, G. (1989) *Genes Dev.*, **3**, 431–437.
- Ma, J. and Ptashne, M. (1987) *Cell*, **48**, 847–853.
- Hope, I.A. and Struhl, K. (1986) *Cell*, **46**, 885–894.
- Hope, I.A., Mahadevan, S. and Struhl, K. (1988) *Nature*, **333**, 635–640.
- Wilusz, J. and Shenk, T. (1990) *Mol. Cell. Biol.*, **10**, 6397–6407.
- Kim, Y.J. and Baker, B.S. (1993) *Mol. Cell. Biol.*, **13**, 174–183.
- Kenan, D.J., Query, C.C. and Keene, J.D. (1991) *Trends Biochem. Sci.*, **16**, 214–220.
- Nietfeld, W., Mentzel, H. and Pieler, T. (1990) *EMBO J.*, **9**, 3699–3705.
- Burd, C.G., Matunis, E.L. and Dreyfuss, G. (1991) *Mol. Cell. Biol.*, **11**, 3419–3424.
- Prieto, J., Candel, E. and Coloma, A. (1991) *Nucl. Acids Res.*, **19**, 1341.
- McDonald, H., Smailus, D., Jenkins, H., Adams, K., Simpson, N.E. and Goodfellow, P.J. (1992) *Genomics*, **13**, 344–348.
- Fukami-Kobayashi, K., Tomoda, S. and Go, M. (1993) *FEBS Letters*, **335**, 289–293.
- Kim, Y.-J., Baker, B.S. (1993) *Mol. Cell. Biol.*, **13**, 174–183.
- Matunis, M.J., Matunis, E.L. and Dreyfuss, G. (1993) *Mol. Cell. Biol.*, **13**, 6114–6123.
- Robinow, S., Campos, A.R., Yao, K.-M. and White, K. (1988) *Science*, **242**, 1570–1572.
- Szabo, A., Dalmau, J., Manley, G., Rosenfeld, M., Wong, E., Henson, J., Posner, J.B. and Furneaux, H.M. (1991) *Cell*, **67**, 325–333.
- Levine, T.D., Gao, F., King, P.H., Andrews, L.G. and Keene, J.D. (1993) *Mol. Cell. Biol.*, **13**, 3494–3504.
- Kim, Y.-J. and Baker, B.S. (1993) *J. Neuroscience*, **13**, 1045–1056.
- Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C. and Luhrmann, R. (1993) *Nature*, **363**, 283–286.
- Cobianchi, F., Calvio, C., Stoppini, M., Buvoili, M., and Riva, S. (1993) *Nucleic Acids Res.*, **21**, 949–955.
- Mayrand, S.H., Dwen, P. and Pederson, T. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7764–7768.
- Chrislip, K.M., Hengst-Zhang, J.A. and Jacob, S.T. (1991) *Gene Expression*, **1**, 197–206.
- Csermely, P., Schnaider, T., Cheatham, B., Olson, M.O.J. and Kahn, C.R. (1993) *J. Biol. Chem.*, **268**, 9747–9752.
- Peter, M., Nakagawa, J., Doree, M., Labbe, J.C. and Nigg, E.A. (1990) *Cell*, **60**, 791–801.
- Paris, J., Swenson, K., Piwnicka-Worms, H. and Richter, J.D. (1991) *Genes Dev.*, **5**, 1697–1708.
- Kwon, Y.K. and Hecht, N.B. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3584–3588.
- Kemp, B.E. and Pearson, R.B. (1990) *Trends Biochem. Sci.*, **15**, 342–346.
- Query, C.C., Bentley, R.C. and Keene, J.D. (1989) *Cell*, **57**, 89–101.