The EMBO Journal vol.14 no.17 pp.4357–4364, 1995

Identification of two KH domain proteins in the α-globin mRNP stability complex

Megerditch Kiledjian[1], Xiaoming Wang and Stephen A.Liebhaber[2]

Howard Hughes Medical Institute, Departments of Genetics and Medicine, University of Pennsylvania School of Medicine, Room 428, Clinical Research Building, 415 Curie Blvd, Philadelphia, PA 19104-6145, USA

[1]Present address: Rutgers University, Department of Biological Sciences, Nelson Biological Laboratories, PO Box 1059, Piscataway, NJ 08855-1059, USA

[2]Corresponding author

Accumulation of globin mRNAs during erythroid differentiation is dependent on their extraordinary stability. The longevity of human α-globin mRNA is associated with a ribonucleoprotein complex (α-complex) formed on the 3′ untranslated region (3′UTR). One or more of the proteins within this α-complex contain strong polycytosine [poly(C)] binding (αPCB) activity. In the present report we purify αPCB activity from human erythroid K562 cells. Although not able to bind the α-globin 3′UTR directly, αPCB activity is sufficient to complement α-complex formation in a cytosolic extract depleted of poly(C) binding activity. Peptide microsequencing demonstrates that αPCB activity contains two structurally related poly(C) binding proteins. These two proteins, α-complex protein (αCP) -1 and -2, have an overall structural identity of 80% and contain three repeats of the K homology (KH) domain which is found in a subset of RNA binding proteins. Epitope-tagged recombinant αCP-1 and αCP-2 expressed in cells are each incorporated into the α-complex. We conclude that αCP-1 and αCP-2, members of the KH domain RNA binding protein family, are involved in formation of a sequence-specific α-globin mRNP complex associated with α-globin mRNA stability. As such this represents the first example of a specific function for this class of proteins and suggests potential roles for other members of this protein family.

Keywords: α-globin mRNA/KH domain/mRNA stability/poly(C) binding/RNA binding proteins/3′ untranslated region

Introduction

The role of post-transcriptional regulation in eukaryotic gene expression is increasingly evident. RNA polymerase II transcripts are extensively processed in the nucleus and subsequently transported to the cytoplasm where they must be maintained in a functional state for efficient recognition as substrate by the translational machinery. Significant progress has been made in defining the biochemistry of the nuclear events in this pathway and in defining corresponding genetic controls (for reviews see Nigg *et al.*, 1991; Wahle and Keller, 1992; Dreyfuss *et al.*, 1993; Lamm and Lamond, 1993; Madhani and Guthrie, 1994). In contrast, relatively little is known regarding the controls, mechanisms and biochemistry which determine the level and functional integrity of cytoplasmic mRNAs. The half-life of a eukaryotic mRNA can be as short as several minutes to as long as several days (Peltz *et al.*, 1991; reviewed in Belasco and Brawerman, 1993). In addition, the stability of a particular species of mRNA can change in response to a wide range of environmental, metabolic and cell cycle events (Atwater *et al.*, 1990; Nielson and Shapiro, 1990; Klausner *et al.*, 1993; Sachs, 1993). As such, post-transcriptional control at the level of mRNA stability has a pivotal role in determining the expression profile of a cell (Ross, 1995).

mRNA stability is determined by a combination of general and specific *cis* elements. Two common structural elements, the 5′ terminal m^7G cap and the 3′ terminal polyadenosine [poly(A)] tail, appear to have central roles in protecting the mRNA from degradation and in transducing changes in stability induced by specific control elements (Furuichi *et al.*, 1977; Sachs, 1993). In addition, a number of *cis*-acting stability determinants specific to a particular species of mRNA or to families of related mRNAs have been identified (Peltz and Jacobson, 1992; Pierrat *et al.*, 1993; Weiss and Liebhaber, 1994). These regulatory elements appear to be most commonly localized to the 3′UTR (Jackson, 1993). In all cases studied, these elements appear to work in conjunction with one or more *trans*-acting factors (Kuhn, 1991; Peltz and Jacobson, 1992, Klausner *et al.*, 1993). How general such mRNA stability control mechanisms are and how the relevant RNA–protein interactions assemble and affect the stability of the substrate mRNA remain undefined.

Globin mRNAs provide an ideal system for the study of mRNA stability determinants. Unlike most mRNAs which have half-lives of several hours, the half-lives of the adult globin mRNAs are measured in days (Lodish and Small, 1976; Volloch and Housman, 1981). Initial insight into the basis of this extraordinary stability has been gained through analysis of naturally occurring genetic mutations at the globin loci in the human population. While a large array of mutations exist that result in the loss of globin gene expression, only a few mutations have been described which specifically affect mRNA stability (Russell and Liebhaber, 1993). One such mutation which has been highly informative in this regard is αConstant Spring (α^{CS}; Clegg *et al.*, 1971; Milner *et al.*, 1971). The α^{CS} mutation happens to be the most common non-deletion α-thalassemia mutation world-wide (Liebhaber, 1989). This single base substitution at the translation termination codon, UAA→CAA, allows the ribosome to read into the 3′UTR for an additional 31 codons before

terminating at an in-frame UAA. Individuals homozygous for this mutation have a severe anemia. While abundant levels of the α^{CS} mRNA are present in transcriptionally active erythrocyte progenitor cells of the bone marrow, this mRNA is undetectable in enucleated circulating reticulocytes (Liebhaber and Kan, 1981; Hunt *et al.*, 1982). This finding suggested that the α^{CS} mutation results in a dramatic destabilization of the α-globin mRNA. Recent studies have directly documented this loss of stability and have demonstrated that it reflects ribosomal read-through into the 3′ untranslated region with consequent disruption of an erythroid-specific mRNA stability determinant (Weiss and Liebhaber, 1994). Subsequent studies have defined three discontinuous C-rich segments within the α-globin 3′UTR as critical *cis*-components of this stability element (Weiss and Liebhaber, 1995). RNA affinity chromatography has identified a set of three cytosolic proteins which bind to these elements in a sequence-specific manner and form an mRNP complex (Wang *et al.*, 1995). One or more of these proteins have strong poly(C) binding activity (αPCB). The presence of this α-complex correlates with the stability of the α-globin mRNA; base substitutions in the 3′UTR which destabilize α-globin mRNA also destabilize the α-complex while mutations in this region which fail to affect α-complex formation fail to have a significant effect on mRNA stability (Wang *et al.*, 1995; Weiss and Liebhaber, 1995).

In the present study we characterize αPCB activity which represents a characteristic component of the α-complex. This activity is purified biochemically and the sequences of two closely related proteins which mediate this activity are determined. We demonstrate that these proteins, each containing three copies of the K homology (KH) domain common to a number of RNA binding proteins, are incorporated into the α-complex and hence are likely to be critical determinants of human α-globin mRNA stability.

Results

αPCB activity is present in K562 cells

We have previously demonstrated that the ability of α-globin mRNA to assemble an RNP complex (α-complex) on the 3′UTR parallels its stability in erythroid cells (Wang *et al.*, 1995; Weiss and Liebhaber, 1995). Furthermore, αPCB activity was found to be an integral component of the α-complex (Wang *et al.*, 1995). In the present report, we purify and characterize αPCB activity from K562 cells, a human erythroid cell line. As an initial step we demonstrate that cytoplasmic S100 extract from K562 cells contains the necessary components for α-complex formation. An RNA gel shift study was carried out in which ^{32}P-labeled α-globin 3′UTR was incubated with K562 cytosolic S100 extract, partially digested with RNase A+T1, and the RNase-resistant complex was electrophoresed on a native polyacrylamide gel (Figure 1). The RNP complex which is formed with K562 extract migrates at the same position as the α-complex previously characterized in MEL cells (Wang *et al.*, 1995) and as would be expected, the formation of this complex is specifically sensitive to competition by poly(C) homoribopolymer. Therefore, K562 cell cytosol contains a full complement of α-complex components including αPCB and can be used as the source of protein for purification of the αPCB activity.

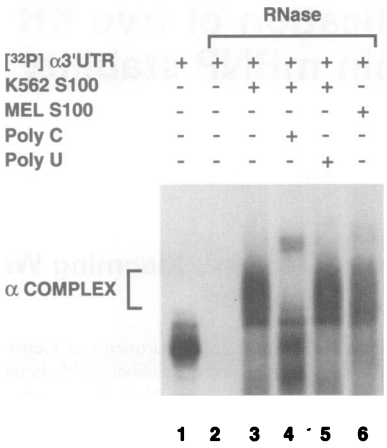

Fig. 1. α-complex and αPCB are present in K562 cells. An electrophoretic mobility shift assay using ^{32}P-internally labeled α-globin mRNA 3′UTR ([^{32}P]α3′UTR) and either K562 cytosolic extract (K562 S100) or MEL cytosolic extract (MEL S100). The first two lanes contain the RNA probe alone in the absence and presence of RNase, respectively. The subsequent lanes are the RNase-resistant α-complex formed on the 3′UTR, in the presence of 200 ng poly(C) competitor RNA (lane 4) or 200 ng poly(U) competitor RNA (lane 5). Migration of the α-complex is indicated.

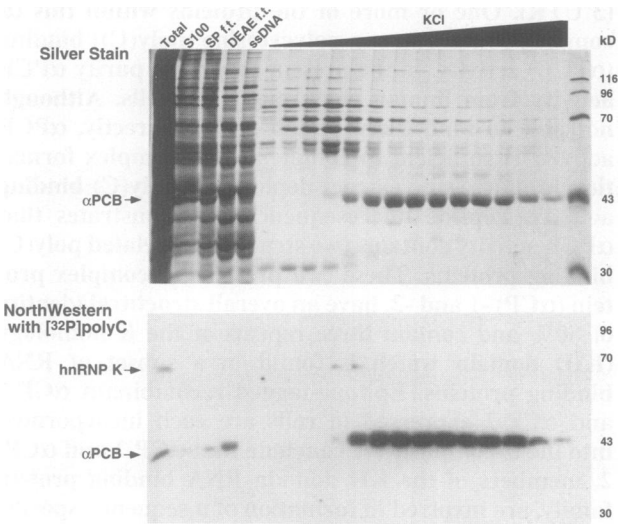

Fig. 2. Purification of the 43 kDa poly(C) binding activity. Fractions from the purification were resolved on a 12.5% SDS–PAGE. The top panel is a silver-stained gel. The first two lanes contain aliquots of K562 cell starting material and cytoplasmic S100 extract, respectively. The next three lanes are the flow-throughs from the SP–Sepharose, DEAE–Sephacel, and single-stranded DNA (ssDNA)–cellulose columns, respectively. The elution profile obtained from an increasing KCl linear gradient (0.3–1.5 M) from the ssDNA–cellulose column is indicated. The bottom panel is a Northwestern analysis of a gel containing aliquots from the same fractions used in the top panel probed with [^{32}P]poly(C). The positions of αPCB and the nuclear poly(C) binding protein, hnRNPK, are indicated.

Purification of αPCB activity from K562 cytosol

αPCB activity was extensively enriched from K562 cytoplasmic S100 fraction by a three-column purification protocol (Figure 2). The presence of the activity was assayed throughout the purification by Northwestern ana-

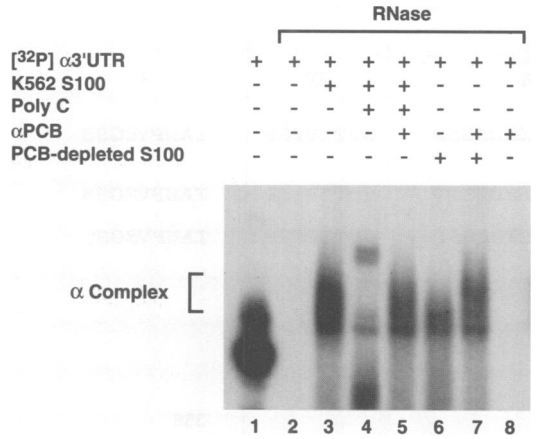

Fig. 3. αPCB is necessary but not sufficient for α-complex formation. αPCB can overcome the effects of poly(C) competition as well as reconstitute the α-complex in a poly(C)-depleted S100 extract. ^{32}P-labeled α-globin 3′UTR mRNA was used in an electrophoretic mobility shift assay with the indicated extracts and competitor RNA as described in the legend to Figure 1. The first two lanes are the 3′UTR RNA probe alone or with RNase, respectively. The RNase-resistant α-complex is shown in the subsequent lanes as indicated. Where specified, 50 ng of competitor poly(C) ribohomopolymer and 50 ng of purified αPCB (αPCB) were used. K562 S100 extract depleted of αPCB activity by adsorption on poly(C)–agarose beads is indicated (PCB-depleted S100).

lysis (see below). The flow-through fractions from DEAE–Sephacel and SP–Sepharose columns contained the entire αPCB activity. This activity was further fractionated by affinity chromatography on single-stranded DNA (ssDNA)–cellulose. As shown on the silver-stained gel in the top panel of Figure 2, a 43 kDa band was enriched during the purification. This band was demonstrated to contain αPCB activity by a Northwestern assay of a parallel gel which was transferred to nitrocellulose membrane and probed with ^{32}P-labeled poly(C) (Figure 2, bottom panel). It should be noted that ssDNA–cellulose fractionation separates the αPCB activity from the 68 kDa nuclear poly(C) binding protein, hnRNP K.

Purified αPCB activity can reconstitute the α-complex

The fractions of the ssDNA affinity column most highly enriched for αPCB were tested for their ability to contribute to the formation of the α-complex. As a first test, poly(C) was added to the S100 extract to inhibit α-complex formation. The αPCB-enriched fraction was then added to this extract and successfully re-established α-complex formation (Figure 3, compare lane 4 with lane 5). This suggested that the activity blocked by the poly(C) competition was present in the enriched αPCB fraction. The experiment was then repeated using an αPCB-depleted extract rather than a poly(C)-competed extract. This was done to address the possibility that αPCB was overcoming poly(C) competition by displacing a distinct poly(C) binding activity from the poly(C) RNA. S100 was depleted of αPCB by repeated adsorptions with poly(C)–agarose beads (see Materials and methods). The depleted extract had a diminished ability to form the α-complex and this activity was significantly enhanced by re-addition of the purified αPCB fraction (Figure 3, compare lanes 6 and 7). These data demonstrate that purified αPCB activity is an integral part of the α-complex. Interestingly, the purified αPCB fraction is unable by itself to form a stable complex with the α-globin 3′UTR (lane 8) and can only do so in conjunction with the non-poly(C) binding components in the S100. Furthermore, it appears that hnRNP K, an avid poly(C) binding protein (Matunis *et al.*, 1992), is not involved in the α-complex since the cytoplasmic S100 is devoid of this nuclear protein (Figure 2) and any trace residual hnRNP K in the fraction should have been removed by the poly(C)–agarose depletion.

αPCB activity consists of at least two distinct KH domain proteins

The αPCB-enriched fractions were pooled and subjected to tryptic protease digestion, peptide isolation and microsequencing. The sequences of two peptides, 1 and 3, were unambiguous while the sequence of the third could be resolved into two distinct peptides (Figure 4A, peptides 2A and 2B). Search of the NCBI BLAST network services (Altschul *et al.*, 1990) revealed sequence identity with two highly similar proteins. Peptides 1 and 3 are found in both proteins while peptides 2A and 2B match sequences specific to each of the two proteins (Figure 4A). Peptide 2A is contained within a human poly(C) binding protein (Figure 5; sub2.3 in Aasheim *et al.*, 1994). Peptide 2B matches perfectly with a sequence within the 362 amino acid murine protein (hnRNP X; Hahm *et al.*, 1993) and an alternative splice form, shorter by 31 amino acids, also containing αPCB activity (mouse C binding protein; Goller *et al.*, 1994). The hnRNP nomenclature was postulated due to the homology noted with hnRNP K. However, there is no evidence to suggest that these proteins are in fact hnRNP proteins (see Discussion). Due to the presence of these two proteins in the mRNP complex associated with α-globin mRNA stability, we suggest that they be referred to as α-complex protein (αCP) -1 and -2 respectively.

Having noted a size discrepancy between the predicted molecular weight of the protein encoded by sub2.3 and the observed size of the αPCB activity as determined by SDS–PAGE, we surmised that these two proteins might represent isoforms possibly generated by a potential alternative splice. To further characterize this difference, we attempted to clone a full-length cDNA corresponding to αCP-1. Towards that end we identified a cDNA clone αCP1-5.1 (Figure 5) by hybridization screening with the sub2.3 cDNA as probe. The sequence of αCP1-5.1 is shown in Figure 5. An additional eight nucleotides within the coding region of αCP1-5.1 were observed compared with the reported sub2.3 sequence (Aasheim *et al.*, 1994). This insertion would result in an extension of the predicted open reading frame by 58 amino acids. To verify this difference we re-sequenced the original sub2.3 clone which revealed that it also contains the same eight nucleotides. The corrected predicted open reading frame of sub2.3 is therefore identical to αCP1-5.1 and both encode a 356 amino acid protein. The extent of the termini present within each clone differs. The 5′ end of αCP1-5.1 extends 92 nucleotides further than sub2.3 and is shorter by 236 nucleotides in the 3′ untranslated region.

A schematic of αCP-1 and αCP-2 is shown in Figure 4B. Both proteins contain three internal peptide repeats corresponding to the KH domain shared by a family of

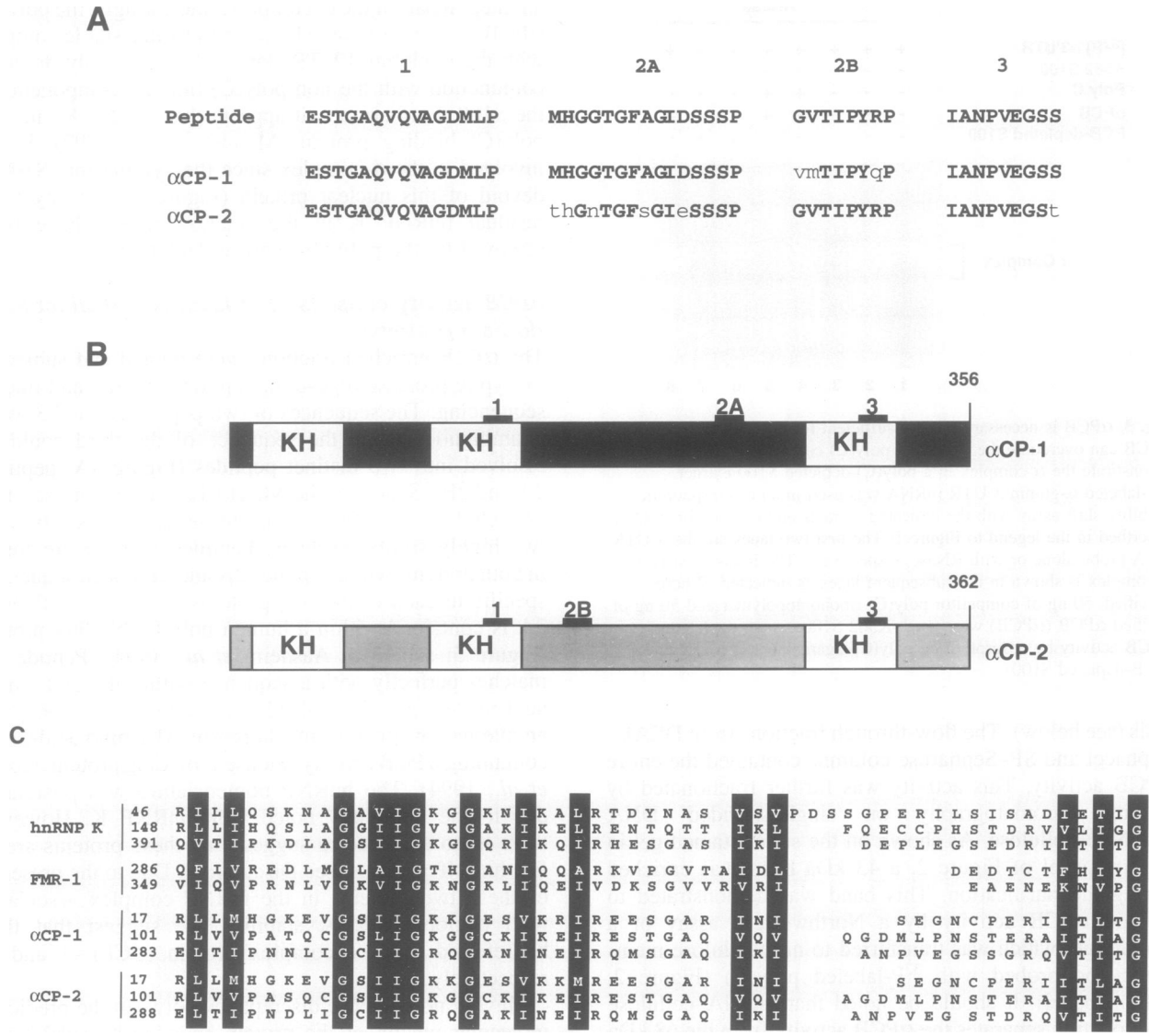

Fig. 4. Identification of αCP-1 and αCP-2. Proteolytic products of purified αPCB activity were microsequenced. (**A**) Amino acids of the sequenced peptides and the designated peptide number are shown on the top line (Peptide). The second and third lines are the corresponding sequences in αCP-1 and αCP-2 (Hahm *et al.*, 1993) respectively. Peptides 1 and 3 contained unique sequences. Peptide 2 contained a mixture of two peptides, 2A and 2B, which correspond to sequences in the αCP-1 and αCP-2 shown below. Upper-case letters are a perfect match to the sequenced peptide, while the lower-case letters denote differences. (**B**) Schematic of αCP-1 and αCP-2 proteins. The positions of the sequenced peptides are indicated by the corresponding peptide number. The KH domains are as indicated. (**C**) An alignment of the KH domains of hnRNP K (Matunis *et al.*, 1992), FMR1 (Verkerk *et al.*, 1991), αCP-1 and αCP-2 (Hahm *et al.*, 1993) is shown. The shaded amino acids are the most highly conserved KH domain consensus sequence. Numbers preceding the peptides indicate the amino acid position in the respective proteins.

RNA binding proteins (Gibson *et al.*, 1993; Burd and Dreyfuss, 1994; Kiledjian *et al.*, 1994). These two proteins have an overall sequence identity of 80% with the strongest homology contained within the three KH domains. An alignment of the three KH domains for human αCP-1 and murine αCP-2 to that of hnRNP K and FMR 1 is shown in Figure 4C. Murine αCP-2 (hnRNP X in Hahm *et al.*, 1993) is identical to the human homolog except for four amino acids in the spacer region between KH2 and KH3 (see below) and was used in the subsequent studies.

αCP-1 and αCP-2 are present in the α-complex

The capacity of recombinant αCP-1 and αCP-2 proteins to incorporate into the α-complex was next tested. A sequence encoding a myc-epitope tag was added to the 5′ terminus of each of the two cDNAs. To ensure that the RNA binding activity is not impaired by the epitope tag, the RNA binding of the fusion proteins was tested. Since both native proteins possess avid αPCB activity (Aasheim *et al.*, 1994; Goller *et al.*, 1994) the binding of the epitope-tagged proteins to poly(C) was tested. Each cDNA was transcribed and translated *in vitro*. ^{35}S-labeled translation products were incubated with poly(C)–agarose or poly(U)–agarose beads. Protein bound to the beads was eluted and resolved on SDS–PAGE. Figure 6 demonstrates that both epitope-tagged proteins retain avid and specific poly(C) binding activity.

The ability of the epitope-tagged proteins to incorporate into the α-complex was next tested. Expression vectors encoding myc-epitope-tagged αCP-1 and αCP-2 proteins were separately transfected into human embryonic kidney 293T cells. Expression in these eukaryotic cells should

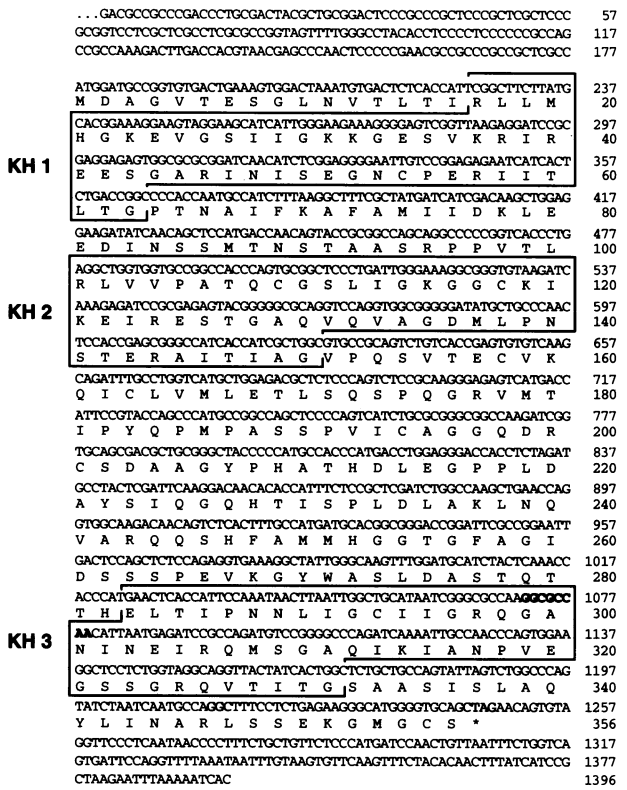

Fig. 5. The nucleotide and predicted open reading frame of αCP1. The nucleotide sequence of clone αCP1-5.1 and the deduced αCP-1 protein are shown. Numbers corresponding to the nucleotide and protein sequence are indicated to the right of each line. The boxes encompass the three KH domains. The bold nucleotides within the third KH domain represent the missing sequences in the reported sub2.3 sequence (Aasheim *et al.*, 1994). (αCP1-5.1, GenBank accession number: U24223).

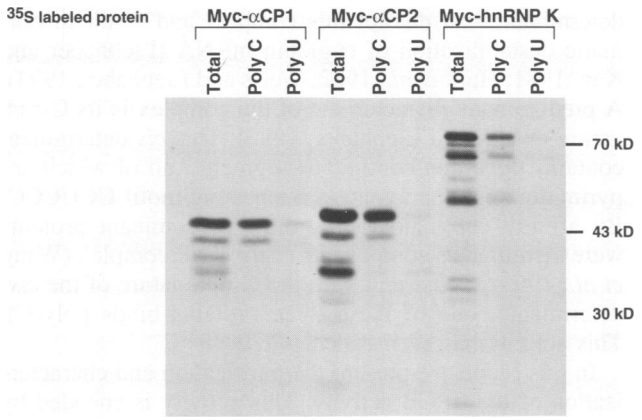

Fig. 6. Poly(C) binding of αCP-1 and αCP2. *In vitro* transcribed and translated [^{35}S]methionine-labeled protein containing the myc-epitope tag was bound to the 30 μg of the indicated ribohomopolymer–agarose beads. An aliquot of the total translation product is shown in the 'Total' lanes, while protein bound to the indicated ribohomopolymer is shown in the Poly C or Poly U lanes. The molecular weight markers are as indicated on the side of the gel.

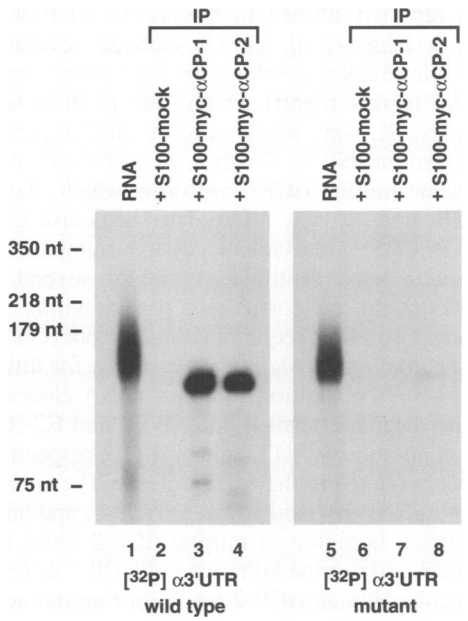

Fig. 7. αCP-1 and αCP-2 are intrinsic to the α-complex. ^{32}P-labeled wild-type α3′UTR (lanes 1–4) or mutant αH19 α3′UTR (lanes 5–8) was used in an RNA binding reaction. RNA was incubated with 293T cell S100 extract from mock-transfected cells (lanes 2 and 6) or cells transfected with plasmid encoding myc-tagged αCP-1 (lanes 3 and 7) or αCP-2 (lanes 4 and 8). The reactions were then subject to immunoprecipitation with a myc-epitope-specific antibody. RNA in the immunoprecipitate was isolated and resolved on a denaturing gel. Lanes 1 and 5 are the wild-type and mutant αH19 probes alone, respectively. The 'IP' above the brackets denotes the lanes containing immunoprecipitation products. The upward tailing in the probe alone lanes could reflect a PCR/*in vitro* transcription artefact. Nucleotide size markers are shown.

allow for any necessary protein modifications which might occur *in vivo*. S100 extract prepared from cells 2 days post-transfection was incubated with ^{32}P-labeled α3′UTR RNA or with an α3′UTR RNA containing a three base substitution which renders it unable to form the α-complex (αH19 in Wang *et al.*, 1995). As demonstrated above (Figure 3), the αCP proteins are unable to bind the α-globin 3′UTR directly and can only do so in the context of the assembled α-complex. Therefore, co-immunoprecipitation of the α-globin 3′UTR RNA with the myc-tagged αCP protein should only occur if myc-αCP is incorporated into the α-complex. The αCP fusion proteins were immunoprecipitated with a myc-tag-specific antibody, 9E10 (Evans *et al.*, 1985). The wild-type α-globin 3′UTR was specifically co-immunoprecipitated from the extracts prepared from the αCP-1- or αCP-2-transfected cells. In contrast, the mutant α-globin 3′UTR was not co-immunoprecipitated in either case (Figure 7, compare lanes 3 and 4 with lanes 7 and 8). The dependence of the co-immunoprecipitation of the labeled RNA on the presence of the myc-tagged transfected αCP proteins is demonstrated by the negative result using untransfected S100 extract (lanes 2 and 6). Therefore, both αCP-1 and αCP-2 are intrinsic to the α-complex.

Discussion

RNA binding proteins play a pivotal role in RNA metabolism and influence the structure and fate of RNA (Kenan *et al.*, 1991; Dreyfuss *et al.*, 1993; Mattaj, 1993; Burd and Dreyfuss, 1994). Stability of the α-globin mRNA and hence its level of expression appears to be determined to a significant extent by the presence of a ribonucleoprotein complex on the 3′UTR (α-complex; Wang *et al.*, 1995). A naturally occurring mutation, α^{CS}, as well as a number of site-directed mutations (Wang *et al.*, 1995) in this

determinant, can disrupt this complex and result in dramatic destabilization of α-globin mRNA (Liebhaber and Kan, 1981; Hunt *et al.*, 1982; Weiss and Liebhaber, 1994). A predominant characteristic of the complex is its C-rich nature (Weiss and Liebhaber, 1995). This *cis*-determinant contains three non-contiguous segments, all of which are pyrimidine pure and contain a repeating motif CCUCCC. By affinity chromatography, three predominant proteins were identified as components of the RNP complex (Wang *et al.*, 1995). Consistent with the C-rich nature of the *cis*-determinant, one of these three proteins binds poly(C). This activity has been called αPCB.

In this report we present the purification and characterization of the αPCB activity. This activity is encoded by at least two closely related proteins, αCP-1 and αCP-2. Although it is clear that both of these proteins are incorporated in the α-complex, it is not established whether they co-exist simultaneously in the complex or whether they are mutually redundant.

The characteristic of αCP-1 and αCP-2 which facilitated their isolation was their αPCB activity. A number of cDNAs encoding poly(C) binding proteins have been recently reported, all of which represent αCP or possible isoforms. Hahm *et al.* (1993) isolated several murine cDNAs (GenBank accession number L19661; referred to as αCP-2 in this report) which, due to their homology with hnRNP K, were assumed to be hnRNP proteins and were named hnRNP X. Goller *et al.* (1994) identified one of the same murine αCP-2 isoforms which, based upon its αPCB activity, was referred to as mouse C binding protein (mCBP; GenBank accession number X75947). This protein was identified as one of several proteins which bound to a cytosine-rich region within a mouse retro-transposon LTR recombination hotspot (Goller *et al.*, 1994). The biological significance, if any, for this binding is unknown. Two additional human cDNA clones submitted to the database, termed hnRNP E1 and E2 (GenBank accession numbers X78137 and X78136, respectively) are also closely related to the αCP proteins. The open reading frame of hnRNP E1 is identical to αCP-1, and hnRNP E2 is the human homolog of murine αCP-2 (99% identity). The human and murine homologs of αCP-2 differ by four amino acids. Human αCP-2 has a four amino acid insert between KH2 and KH3 (at amino acid 168), not present in murine αCP-2. The hnRNP nomenclature which has been applied to these proteins appears to be a misnomer, since the majority of the αPCB activity is in the cytoplasm (Wang *et al.*, 1995) and we are unable to detect epitope-tagged αCP-1 or αCP-2 protein products in the hnRNP complex (data not shown). In addition, a putative alternatively spliced form of αCP-1 reported by Aasheim *et al.* (1994; sub2.3) appears to contain a sequencing error and the corrected open reading frame of sub2.3 is in fact identical to αCP-1 (see above) and hnRNP E1.

The most notable structural characteristic of the αCP proteins is the triple repeat of the KH domain. The KH domain is the newest member of the putative RNA binding motifs and was originally recognized as a repeated motif in hnRNP K for which it was named (K-Homology; Siomi *et al.*, 1993a). It is ~50 amino acids long and highly conserved throughout evolution. Several KH-containing proteins, most notably hnRNP K and FMR1, can bind RNA *in vitro* (Matunis *et al.*, 1992; Wong *et al.*, 1992; Ashley *et al.*, 1993; Siomi *et al.*, 1993b, 1994). A particularly informative mutation in the FMR1 gene suggests that RNA binding may be an important attribute of these proteins. A single point mutation of a conserved residue in one of the KH domains of the FMR1 gene product resulted in severe X-linked mental retardation in a patient (De Boulle *et al.*, 1993) and this same mutation impairs RNA binding *in vitro* (Siomi *et al.*, 1994). However, identification of the cognate RNA bound by FMR1 or any of the other KH domain proteins has thus far proven elusive. Although the KH domain is essential for RNA binding, it is not clear whether this is the region which makes direct contact with RNA. The recent demonstration that the KH domain folds into a structure containing three anti-parallel β-sheets and two α-helices (Morelli *et al.*, 1995) suggests that the β-sheets might provide a binding platform for RNA similar to the four β-sheets of the small nuclear ribonucleoprotein U1A and the hnRNP C protein RNA binding domains (reviewed in Kiledjian *et al.*, 1994; Nagai and Mattaj, 1994). The involvement of the KH domain in additional functions such as protein–protein interactions can not be ruled out.

Northwestern analysis of whole-cell extracts with [^{32}P]poly(C) reveals two major αPCB activities (Wang *et al.*, 1995). One of these at 68 kDa is hnRNP K, while the second band of activity at 43 kDa contains the two co-migrating αCP proteins (this study). Both the hnRNP K and the αCP proteins have three repeats of the KH domain. How the common KH domain structure relates to this common characteristic of poly(C) binding is not clear. The importance, if any, of the αPCB activity *per se* also remains to be established. In the case of αCP-1 and αCP-2, the relevance of poly(C) binding to its activity in α-complex formation is particularly perplexing. The αCP proteins have αPCB activity and they contribute to the α-complex RNP. However, these proteins are not able to bind to the α-3′UTR *cis* element in the absence of the additional α-complex proteins [which are themselves not poly(C) binding] even though this *cis* element is C-rich. Thus the importance of αPCB activity of αCP-1 and αCP-2 to their function(s) and how this activity relates to the formation and function of the α-complex remains to be determined.

The αCP proteins contribute to α-complex formation, which in turn appears to be involved in the establishment of α-globin mRNA stability. Northern blot analysis indicates the mRNA encoding αCP-1 is present in numerous cell types and particularly enriched in skeletal muscle, thymus and peripheral blood leukocytes (Aasheim *et al.*, 1994). It is apparent therefore that α-globin mRNA stability is not likely to be the sole function of these proteins. The α-complex can in fact be formed from the cytosolic S100 extracts of a variety of cells, erythroid and non-erythroid (Wang *et al.*, 1995). The lack of direct αCP binding to the α-globin 3′UTR suggests that protein–protein interactions are required. It will be interesting to determine if the same set of proteins, or a combination thereof, bind to and protect other mRNAs. It is also possible that αCP could interact with additional protein(s) to further increase the repertoire of RNAs to which it can bind and on which it can exert an influence. Identification of the proteins with which the αCP proteins interact in

erythroid and non-erythroid cells will begin addressing their possible additional functions.

Materials and methods

Plasmid construction and isolation of cDNA clone

The insert for the pMyc–αCP-1 plasmid was generated by PCR amplification of αCP1-5.1 clone which contains the full-length coding region of αCP-1. The 5′ primer (5′-AAGCCGAATTCCATGG-ATGCCGGTGTGACTG-3′) introduces an *Eco*RI site just upstream of the first methionine and the 3′ primer (5′-GGAACCTCGAGTGTTCTA-GCTGCACCCCATG-3′) introduces an *Xho*I site 3′ of the stop codon. The amplified fragment was inserted into the *Eco*RI and *Xho*I sites of the pcDNA1–myc plasmid (kindly provided by Matthew Michael, University of Pennsylvania) in frame and downstream of the 11-residue myc epitope tag. The pMyc–αCP-2 plasmid was constructed similarly. The 5′ primer (5′-TAACTGAATTCCATGGACACCGGTGTGATTG-3′) and the 3′ primer (5′-GTCTGCTCGAGTCTAGCTGCTCCCC-ATGCC-3′) were used to amplify the αCP-2 coding region from the hnRNP X cDNA reported in (Hahm *et al.*, 1993). The pMyc–αCP-1 and pMyc–αCP-2 plasmids can be transcribed from the T7 promoter *in vitro* or from the cytomegalovirus promoter in transfected eukaryotic cells. The encoded αCP-1 and αCP-2 proteins contain a myc-epitope tag at their amino termini. All plasmid constructions were confirmed by sequencing. HeLa D98 λgt11 cDNA library (kindly provided by Paula Henthorn and Dave Lukac, University of Pennsylvania) was screened by hybridization screening using the sub2.3 cDNA (Aasheim *et al.*, 1994) using standard techniques (Maniatis *et al.*, 1982).

Cell culture

Human embryonic erythroleukemia K562 cells were grown in RPMI 1640 supplemented with 10% bovine calf serum containing 100 U/ml penicillin and 100 μg/ml streptomycin. Human embryonic kidney 293T cells (kindly provided by M.Malim, University of Pennsylvania) were grown in DMEM supplemented with 10% FCS, 800 μg/ml G418, 100 U/ml penicillin and 100 μg/ml streptomycin. Transfections of 293T cells were carried out with 20 μg of the indicated plasmid using calcium phosphate co-precipitation. Cells were harvested and proteins isolated 2 days post-transfection as described below.

Protein extract preparation

S100 extracts were prepared from K562 or 293T cells as described (Wang *et al.*, 1995). Cells were pelleted, washed twice in PBS and re-suspended in ice-cold lysis buffer (20 mM HEPES, pH 7.6, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 2 μg/ml leupeptin and 0.5% aprotinin) at 5×10^7 cells/ml. All subsequent steps were carried out at 4°C. Cells were lysed with 25 strokes of the type B pestle in a Dounce homogenizer, and the released nuclei and intact cells were pelleted at 2000 *g* for 5 min. The KCl concentration of the supernatant was adjusted to 150 mM and centrifuged at 100 000 *g* for 60 min. The supernatant (S100) was collected, glycerol was added to 5% (v/v), and the extract was adjusted to 5–10 μg/ml with the Centriprep 10 concentrator (Amicon). Protein concentrations were determined by the Bio-Rad protein assay using a bovine serum albumin standard curve.

Poly(C)-depleted S100 extracts were prepared by incubation of the S100 with poly(C)–agarose beads at 4°C for 15 min in HEG (50 mM HEPES, 1 mM EDTA and 10% glycerol) buffer with 300 mM KCl. Unbound protein was recovered and poly(C) binding was repeated twice more before use in the electrophoretic mobility shift assays.

Electrophoretic mobility shift assay

RNA binding and electrophoretic mobility shift assays were carried out as described by Wang *et al.* (1995) with minor modifications. Approximately 10 000 c.p.m. (0.5 ng) of *in vitro*-transcribed ^{32}P uniformly labeled RNA probe was used per reaction. The RNA was denatured at 90°C for 3 min followed by 20 min incubation at room temperature before use. Reactions were carried out in a 15 μl total volume in 10 mM Tris–HCl, pH 7.4, 1.5 mM $MgCl_2$, 150 mM KCl, 25 μM *p*-nitophenyl-*p*-guanidinobenzoate and 10 μg/ml leupeptin. S100 was pre-treated for 15 min on ice with Inhibitor-ACE (1 U per 30 μl S100 extract, 5′–3′) and 1% β-mercaptoethanol. The RNA probe was incubated with 40 μg of the treated S100 for 20 min at room temperature. Unbound RNA was degraded with 1 U RNase T1 (Boehringer Mannheim) and 10 ng RNase A (Sigma) for 10 min at room temperature followed by addition of heparin to a final concentration of 5 mg/ml and incubated for an additional 10 min. Where indicated, non-radioactive poly(C) or poly(U) (Sigma) competitor RNA was mixed with ^{32}P-labeled probe RNA before incubation with the S100. The binding reactions were run on a 5% polyacrylamide gel (60:1 acrylamide:bis) in 0.5× TBE buffer at 8 V/cm. The gel was dried and exposed onto Kodak X-ray film.

Protein purification

Cytoplasmic S100 extract was isolated from 10^{10} K562 cells as described above except the KCl concentration was adjusted to 40 mM. The S100 was treated with micrococcal nuclease (400 U/ml) in the presence of 1 mM $CaCl_2$ for 20 min at 30°C followed by the addition of 5 mM EGTA. Insoluble protein was pelleted at 10 000 *g* for 10 min and the supernatant was loaded onto a column containing 40 ml of SP–Sepharose (Pharmacia) pre-washed in 1 M KCl and equilibrated in HEG (40 mM KCl). The flow through was collected and passed through a column containing 40 ml of DEAE–Sephacel (Pharmacia) which had been pre-washed and equilibrated as for the SP–Sepharose column. The KCl concentration of the DEAE–Sephacel flow through was adjusted to 200 mM and bound to ssDNA–cellulose, pre-washed in 2 M KCl buffer, and equilibrated in HEG (200 mM KCl). The ssDNA–cellulose column was washed with heparin (1 mg/ml) followed by HEG (300 mM KCl). Bound protein was eluted with a linear KCl gradient (0.3–1.5 M) over 90 min on the WATERS 650 Protein Purification system (Millipore). Aliquots (20 μl) from the 1 ml fractions were mixed with an equal volume of 2× SDS-loading buffer and resolved on 12.5% SDS–PAGE.

Peptide microsequencing

Purified αPCB activity was subject to tryptic digestion (2% trypsin in 0.1 M NH_4HCO_3, pH 8.1) at 37°C overnight. Proteolytic peptides were purified by HPLC on a reverse-phase C4 column. Individual peptide peaks were subjected to an Applied Biosystems 473A gas phase microsequencer (Medical Center Protein Chemistry Facility of the University of Pennsylvania). Peptide 2 was a mixture of two peptides that could be separately aligned with αCP-1 and αCP-2. The specificity of trypsin for the lysine at the carboxyl end of Peptide 2B (in αCP-2) allowed for the distinction of this peptide sequence from that of Peptide 2A.

Northwestern analysis

Proteins separated on 12.5% SDS–PAGE were electroblotted onto nitrocellulose for 1 h at 150 mA in transfer buffer (50 mM Tris, 14.5 mM glycine and 20% methanol). Northwestern (NW) blot analysis with ^{32}P-5′-end-labeled poly(C) was carried as out as described in Wang *et al.* (1995).

Ribonucleotide homopolymer binding assay

Binding of *in vitro* produced protein was carried out as described in Kiledjian and Dreyfuss (1992). A total of 10^5 c.p.m. of trichloroacetic acid-precipitable counts of *in vitro*-produced protein was bound to the 30 μg of the indicated ribonucleotide homopolymer–agarose beads (Sigma) in binding buffer [10 mM Tris–HCl, 250 mM NaCl, 1.5 mM $MgCl_2$, 0.5% Triton X-100 (v/v), 2 μg/ml leupeptin and pepstatin and 0.5% aprotinin] for 15 min at 4°C. Protein bound to the ribonucleotide homopolymer–agarose beads were pelleted by brief centrifugation and washed for an additional 10 min in binding buffer containing 500 mM NaCl and 1 mg/ml heparin. The beads were subsequently washed five times in binding buffer. The beads were re-suspended in SDS–PAGE loading buffer and boiled for 3 min to elute bound protein from the nucleic acid. The eluted protein was resolved on 12.5% SDS–PAGE and visualized by fluorography.

Immunoprecipitations

S100 from 293T cells transfected with 20 μg pMyc–αCP-1 or pMyc–αCP-2 was used in a modified RNA binding assay. 100 μg of S100 pre-treated with Inhibitor-ACE (1 U per 30 μl S100 extract, 5′–3′) and 1% β-mercaptoethanol for 15 min on ice was incubated with 100 000 c.p.m. of *in vitro*-generated [^{32}P]α3′UTR RNA (see above) or mutant [^{32}P]α3′UTR RNA (αH19 in Wang *et al.*, 1995). Binding reactions were carried out at room temperature for 20 min followed by a 10 min incubation with 2.5 mg/ml heparin. Immunoprecipitation of the myc-tagged protein and accompanying complex was carried out on a rocking platform at 4°C for 30 min with 4 μl of 9E10 antibody ascites fluid. Immune complexes were bound to *Staphylococcus aureus* protein A–Sepharose beads (Pharmacia) which were then pelleted by brief centrifugation, washed by rocking at 4°C for 10 min with PBS containing 0.1% Triton X-100 (v/v) and 1 mg/ml heparin, and then rinsed five times with the same buffer and once with PBS. The RNA–protein complex was

disrupted by boiling for 3 min in 50 μl TE with 1% SDS. The RNA was ethanol-precipitated following phenol–chloroform extractions, resuspended in 80% formamide, resolved on a 7 M urea/6% polyacrylamide gel and the [^{32}P]RNA visualized by autoradiography.

Acknowledgements

We would like to thank S.Smale for providing the hnRNP X clone, H.-C.Aasheim for the sub2.3 clone, the National Cell Culture Center (Cellex Biosciences, Inc.) for assistance in growing K562 cells, and Hong Lam for technical support in the library screen. We would also like to thank Michael Malim, Tom Kadesch, Bill Lee and members of the Liebhaber laboratory for helpful discussions and critical reading of the manuscript. This work was partially supported by NIH grant P60-HL38632 to S.A.L. who is an Investigator in the Howard Hughes Medical Institute.

References

Aasheim,H.-C., Loukianova,T., Deggerdel,A. and Smeland,E.B. (1994) Tissue specific expression and cDNA structure of a human transcript encoding a nucleic acid binding [oligo(dC)] protein related to the pre-mRNA binding protein K. *Nucleic Acids Res.*, **22**, 959–964.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Ashley,C.T.,Jr, Wilkinson,K.D., Reines,D. and Warren,S.T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Science*, **262**, 563–566.

Atwater,J.A., Wisdom,R. and Verma,I.M. (1990) Regulated mRNA stability. *Annu. Rev. Genet.*, **24**, 519–641.

Belasco,J. and Brawerman,G. (1993) *Control of mRNA Stability.* Academic Press, San Diego, CA.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structure and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Clegg,J.B., Weatherall,D.J. and Milner,P.F. (1971) Haemoglobin Constant Spring – α chain termination mutant? *Nature*, **234**, 337–340.

De Boulle,K. *et al.* (1993) Point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nature Genet.*, **3**, 31–35.

Dreyfuss,G., Matunis,M.J., Piñol-Roma,S. and Burd,C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.*, **62**, 289–321.

Evans,G.I., Lewis,G.K., Ramsay,G. and Bishop,J.M. (1985) Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.*, **5**, 3610–3616.

Furuichi,Y., LaFiandra,A. and Shatkin,A.J. (1977) 5′-Terminal structure and mRNA stability. *Nature*, **266**, 235–239.

Gibson,T.J., Thompson,J.D. and Heringa,J. (1993) The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding nucleic acid. *FEBS Lett.*, **324**, 361–366.

Goller,M., Funke,B., Gehe-Becker,C., Kröger,B., Lottspeich,F. and Horak,I. (1994) Murine protein which binds preferentially to oligo-C-rich single stranded nucleic acids. *Nucleic Acids Res.*, **22**, 1885–1889.

Hahm,K., Kim,G., Turck,C.W. and Smale,S.T. (1993) Isolation of a murine gene encoding a nucleic acid-binding protein with homology to hnRNP K. *Nucleic Acids Res.*, **21**, 3894.

Hunt,D.M., Higgs,D.R., Winichagoon,P., Clegg,J.B. and Weatherall,D.J. (1982) Haemoglobin Constant Spring has an unstable α chain messenger RNA. *Br. J. Haematol.*, **51**, 405–413.

Jackson,R.J. (1993) Cytoplasmic regulation of mRNA function: the importance of the 3′ untranslated region. *Cell*, **74**, 9–14.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.

Kiledjian,M. and Dreyfuss,G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.*, **11**, 2655–2664.

Kiledjian,M., Burd,C.G., Görlach,M. Portman,D.S. and Dreyfuss,G. (1994) Structure and function of hnRNP proteins. In Nagai,K. and Mattaj,I.W. (eds), *RNA–Protein Interactions.* IRL Press, Oxford, pp. 127–149.

Klausner,R.D., Rouault,T.A. and Harford,J.B. (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell*, **72**, 19–28.

Kuhn,L.C. (1991) mRNA–protein interactions regulate critical pathways in cellular iron metabolism. *Br. J. Haematol.*, **79**, 1–5.

Lamm,G.M. and Lamond,A.I. (1993) Non-snRNP protein splicing factors. *Biochim. Biophys. Acta.*, **1173**, 247–265.

Liebhaber,S.A. (1989) Alpha thalassemia. *Hemoglobin*, **13**, 685–731.

Liebhaber,S.A. and Kan,Y.W. (1981) Differentiation of the mRNA transcripts originating from the α1- and α2-globin loci in normals and α-thalassemics. *J. Clin. Invest.*, **68**, 439–446.

Lodish,H.F. and Small,B. (1976) Different lifetimes of reticulocyte messenger RNA. *Cell*, **7**, 59–65.

Madhani,H.D. and Guthrie,C. (1994) Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1–26.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mattaj,I.W. (1993) RNA recognition: a family matter? *Cell*, **73**, 837–840.

Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1992) Characterization and primary structure of the poly(C)-binding hnRNP K protein. *Mol. Cell. Biol.*, **12**, 164–171.

Milner,P.F., Clegg,J.B. and Weatherall,D.J. (1971) Haemoglobin-H disease due to a unique haemoglobin variant with an elongated α-chain. *Lancet*, **10**, 729–732.

Morelli,M.A.C., Stier,G., Gibson,T., Joseph,C., Musco,G., Pastore,A. and Travè,G. (1995) The KH module has an αβ fold. *FEBS Lett.*, **358**, 193–198.

Nagai,K. and Mattaj,I.W. (1994) RNA–protein interactions in the splicing snRNPs. In Nagai,K. and Mattaj,I.W. (eds), *RNA–Protein Interactions.* IRL Press, Oxford, pp. 151–177.

Nielson,D.A. and Shapiro,D.J. (1990) Insights into hormonal control of messenger RNA stability. *Mol. Endocrinol.*, **4**, 953–957.

Nigg,E.A., Baeuerle,P.A. and Lührmann,R.A. (1991) Nuclear import-export: in search of signals and mechanisms. *Cell*, **66**, 15–22.

Peltz,S.W. and Jacobson,A. (1992) mRNA stability: in *trans*-it. *Curr. Opin. Cell Biol.*, **4**, 979–983.

Peltz,S.W., Brewer,G., Bernstein,P., Hartford,P.A. and Ross,J. (1991) Regulation of mRNA turnover in eucaryotic cells. *Crit. Rev. Eucaryot. Gene Expr.*, **1**, 99–126.

Pierrat,B., Lacroute,F. and Losson,R. (1993) The 5′ untranslated region of the PRP1 regulatory gene dictates rapid mRNA decay in yeast. *Gene*, **131**, 43–51.

Ross,J. (1995) mRNA stability in mammalian cells. In Joklik,W.K. (ed.), *Microbiology Reviews*, in press.

Russell,J.E. and Liebhaber,S.A. (1993) Molecular genetics of thalassemia. *Adv. Genome Biol.*, **2**, 283–353.

Sachs,A.B. (1993) Messenger RNA degradation in eukaryotes. *Cell*, **74**, 413–421.

Siomi,H., Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1993a) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.

Siomi,H., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1993b) The protein product of the Fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell*, **74**, 291–298.

Siomi,H., Choi,M., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domains of FMR1 that causes fragile X syndrome. *Cell*, **77**, 33–39.

Verkerk,A.J. *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.

Volloch,V. and Housman,D. (1981) Stability of globin mRNA in terminally differentiating murine erythroleukemia cells. *Cell*, **23**, 509–514.

Wahle,E. and Keller,W. (1992) The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.*, **61**, 419–440.

Wang,X., Kiledjian,M., Weiss,I. and Liebhaber,S.A. (1995) Detection and characterization of a 3′ untranslated region ribonucleoprotein complex associated with human α-globin mRNA. *Mol. Cell. Biol.*, **15**, 1769–1777.

Weiss,I.M. and Liebhaber,S.A. (1994) Erythroid cell-specific determinants of alpha-globin mRNA stability. *Mol. Cell. Biol.*, **14**, 8123–8132.

Weiss,I.M. and Liebhaber,S.A. (1995) Erythroid cell-specific mRNA stability elements in the α2-globin 3′ nontranslated region. *Mol. Cell. Biol.*, **15**, 2457–2465.

Wong,G., Müller,O., Clark,R., Conroy,L., Moran,M.F., Polakis,P. and McCormick,F. (1992) Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell*, **69**, 551–558.

Received on April 24, 1995; revised on June 12, 1995