Biochemistry

 Volume 32, Number 12 March 30, 1993

Perspectives in Biochemistry

Polypyrimidine Tracts and Their Binding Proteins: Regulatory Sites for Posttranscriptional Modulation of Gene Expression†

David R. Morris,*,‡ Tomohito Kakegawa,‡ Roger L. Kaspar,‡ and Michael W. White§

Department of Biochemistry, SJ-70, University of Washington, Seattle, Washington 98195, and Department of Veterinary Molecular Biology, Montana State University, Bozeman, Montana 59717

Received November 23, 1992

Extended homopyrimidine or pyrimidine-rich sequences are present in functional regions of the 5′-leaders and the 3′-untranslated regions (3′-UTRs) of mRNA molecules and within introns of pre-mRNAs. These polypyrimidine (pPy) tracts are now emerging as important motifs in a variety of aspects of RNA metabolism, including regulation of translation of specific mRNA species, pre-mRNA splicing, cytoplasmic degradation and polyadenylation of mRNA, and transcriptional termination. In some of these instances, proteins interacting with these pPy tracts have been identified, and their roles in regulation of gene expression are beginning to be understood. The purpose of this Perspective is to review the occurrence of pPy tracts in various RNA species and to discuss the roles of these sequences and their binding proteins in posttranscriptional processes.

Translational Control of Ribosomal Protein Synthesis

The synthesis of ribosomal proteins (r-proteins) has been shown to be under translational control [reviewed in Kaspar et al. (1993)] in a variety of eukaryotic cells (DePhilip et al., 1980; Geyer et al., 1982; Tushinski & Warner, 1982; Meyuhas et al., 1987; Agrawal & Bowman, 1987; Amaldi et al., 1989; Kaspar et al., 1990, 1992). In each case regulation involves movement of r-protein mRNAs between untranslated subribosomal particles and active polysomes. Regulation of this process is characterized by a distinct bimodal distribution of r-protein mRNA in the cytosol, i.e., the mRNA is either located in mRNP particles or fully loaded with ribosomes, indicating that once an r-protein mRNA enters a translatable pool, ribosomal initiation is quite efficient. Because translational control of the synthesis of the r-proteins is coordinate (Kaspar et al., 1993), it seems likely that there is a common mechanism of regulation for the roughly 70 different species of r-protein mRNAs.

The 5′-transcript leaders (5′-TLs) of several r-protein mRNAs have been found to confer translational regulation on an unregulated reporter mRNA (Mariottini & Amaldi, 1990; Levy et al., 1991; Hammond et al., 1991). Wagner and Perry (1985) first noted the existence of pPy tracts that spanned the site of transcriptional initiation of r-protein genes; sequences of 8–14 consecutive pyrimidine residues have been found invariably in the 5′-TLs of all cloned vertebrate r-protein mRNAs [see Kaspar et al. (1993)]. Deletion of the pPy region (Kaspar et al., 1992) or substituting purines for pyrimidines (Levy et al., 1991; our unpublished results) destroyed the ability of the altered mRNAs to be sequestered into mRNP particles and rendered their translation independent of growth state. It is interesting to note that a number of translationally regulated mRNAs, not encoding r-proteins, have been found to share the common feature of pPy tracts in their 5′-TLs (Pinsky et al., 1985; Chitpatima et al., 1988; Makrides et al., 1988; Kerfelec et al., 1990). Although functional roles for the pPy sequences in these mRNAs have not yet been demonstrated, by analogy to the role of pPy tracts in r-protein messages, one suspects that these sequences may be a commonly used element for translational regulation in vertebrate cells.

A 56-kDa protein ($p56^{L32}$) was cross-linked specifically in vitro to RNA containing the r-protein L32 pPy element (Kaspar et al., 1992) and did not bind to RNA in which the pPy tract had been deleted (Kaspar et al., 1992) or mutated (our unpublished results). The binding activity of $p56^{L32}$ did

† The studies from our laboratories were supported by research grants to D.R.M. (CA39053) and M.W.W. (GM42791) and a predoctoral traineeship to R.L.K. (HD07183).

* Please address correspondence to this author.
‡ University of Washington.
§ Montana State University.

0006-2960/93/0432-2931$04.00/0

not seem to be dependent on the growth state of the cells from which it was extracted (Kaspar et al., 1992), suggesting that interaction of $p56^{L32}$ with RNA is not the regulated step in translational repression in this system. Although an intact pPy tract is clearly necessary for $p56^{L32}$ binding, competition experiments suggest that other regions of the 5′-TL may be important as well (our unpublished results). The minimal nucleotide sequences necessary for $p56^{L32}$ binding are not known, and it also remains to be determined whether $p56^{L32}$ will interact with r-protein mRNAs other than that encoding L32. Although $p56^{L32}$ shows some similarities to two pPy binding proteins (pPyBPs) thought to be involved in splicing, evidence to date suggests that they are distinct proteins (see below).

Processing of mRNA Precursors

Removal of introns from pre-mRNA molecules proceeds through formation of an intermediate complex known as the spliceosome (Grabowski & Sharp, 1986). This structure contains the 5′ and 3′ splice sites of the pre-mRNA bound to snRNPs and other components of the active splicing complex. The recognition processes underlying assembly of this structure are complex [for reviews, see Padgett et al. (1986), Sharp (1987), and Krainer and Maniatis (1988)] but involve, in part, a pPy sequence located between the branch site and the 3′ splice site in introns of higher eukaryotes (Frendewey & Keller, 1985; Ruskin & Green, 1985; Reed & Maniatis, 1985) and yeast (Patterson & Guthrie, 1991). The composition and position of the pPy tract, as well as the sequence of the branchpoint, determine the strength of selection of a particular 3′ splice site (Freyer et al., 1989; Reed, 1989; Smith et al., 1989; Dominski & Kole, 1992). The pPy tract seems to be involved at an early step in assembly of an active splicing complex and is required for formation of the intermediate lariat structure (Frendeway & Keller, 1985; Reed & Maniatis, 1985; Ruskin & Green, 1985; Binderelf & Green, 1987; Freyer et al., 1989; Reed, 1989).

Interactions at the pPy site may be of regulatory significance in the choice of alternative 3′ (Mullen et al., 1991; Patterson & Guthrie, 1991; Dominski & Kole, 1992) and 5′ (Fu et al., 1988) splice sites. For example, the "default" mode of splicing the mutually exclusive exons 2 and 3 of the α-tropomyosin transcript is determined by the lengths and pyrimidine contents of the pPy tracts of the competing 3′ splice sites (Mullen et al., 1991). The product of the *sex-lethal* gene in *Drosophila* binds to the pPy region at the 3′ splice site of several pre-mRNAs and may thereby regulate alternative splicing (Inoue et al., 1990), perhaps by competition with splicing factor U2AF [discussed in Frankel et al. (1991)]. Also in *Drosophila*, the pPy tract of the female-specific 3′ splice site of the *doublesex* gene contains intervening purine residues and mutation of the purines to pyrimidines causes constitutive utilization of the female-specific site (Hoshijima et al., 1991). The products of the *transformer* and *transformer-2* genes activate splicing at the wild-type female site of *doublesex* through interaction with an exon sequence, perhaps by promoting assembly of a nucleoprotein complex at the adjacent weak 3′ splice site (Tian & Maniatis, 1992).

Several groups have identified proteins that interact with pPy sequences at 5′ splice sites. A 62-kDa pPyBP (pPTB) in nuclear extracts was cross-linked to radioactive intron sequences by irradiation with UV light (Garcia-Blanco et al., 1989; Wang & Pederson, 1990). Using a different approach, Zamore and Green (1989) identified a factor of 65 kDa, U2AF, that complemented the splicing reaction in vitro. Binding of U2AF to the pPy tract is required for assembly of an active spliceosomal complex. Although the apparent molecular weights of pPTB and U2AF are similar, and both bind to pPy sequences, the two proteins seem quite different, based on purification properties, on the inability of pPTB to replace U2AF in splicing reactions (Zamore & Green, 1991), and on comparison of cDNA clones (see below). It should be noted that, although the ability of U2AF to complement in vitro splicing reactions has been clearly demonstrated, no physiological function of pPTB has yet been established. It is interesting, however, that an isoform of pPTB has been shown to be associated with hnRNP particles and to be identical to hnRNP I (Ghetti et al., 1992).

A U2AF cDNA encodes a protein of 53 kDa that contains sequences predicted from microsequencing of material purified from cell extracts (Zamore et al., 1992). The calculated molecular mass contrasts with the apparent molecular mass of 63 kDa determined by SDS gel electrophoresis. The protein encoded by the cDNA clone contained two motifs found frequently in splicing factors: a region rich in arginine and serine (RS) and three copies of the ribonucleoprotein consensus sequence (RNP-CS). The RS domain in U2AF is essential for activity, but deletion of this sequence did not interfere with RNA-binding activity (Zamore et al., 1992). Mutation studies demonstrated that all three RNP-CS domains were necessary for high-affinity binding to RNAs containing pPy tracts, as well as for splicing activity (Zamore et al., 1992). A reasonable model is that the RNP-CS domains of U2AF are responsible for sequence-specific binding to the branch points of introns and that the RS domain is involved in other interactions governing spliceosome assembly and/or splicing.

Clones encoding pPTB have been isolated from human (Gil et al., 1991; Patton et al., 1991), rat (Brunel et al., 1991), and mouse (Bothwell et al., 1991) cDNA libraries. All encode proteins of approximately 57 kDa in size. The RNA-binding domain has been localized to the carboxy-terminal 195 amino acids of murine pPTB (Bothwell et al., 1991). Although pPTB shows some similarity to other RNA-binding proteins along its length, there is no strong homology to the RNP-CS domains that have been found to be responsible for binding of U2AF and other proteins to RNA (Gil et al., 1991; Patton et al., 1991; Bothwell et al., 1991). Interestingly, multiple cDNAs of human origin have been isolated that probably arose from alternatively spliced transcripts and seem to have the potential of producing isoforms of pPTB (Gil et al., 1991). These alternative amino acid sequences occur to the immediate amino-terminal side of the RNA-binding region defined by Bothwell et al. (1991). It is interesting that pPTB and U2AF, which bind to similar sequences at 3′ splice sites, seem to have quite different structural determinants in their RNA-binding domains.

Although a systematic comparison of the three proteins has not been carried out, U2AF, pPTB, and r-protein regulatory factor $p56^{L32}$ seem to share many properties in common. The three proteins have apparent molecular sizes in the range 56–65 kDa. The binding sites for these proteins all contain pPy tracts, with no apparent consensus sequence or secondary structure. pPTB and $p56^{L32}$ both bind single-stranded DNA with the same specificity as RNA (Wang & Pederson, 1990; Bothwell et al., 1991; our unpublished results). This latter activity accounts for the fact that pPTB has been isolated and cloned using binding to oligodeoxynucleotide probes as an assay (Brunel et al., 1992; Jansen-Durr et al., 1992). Similar to U2AF (Zamore & Green, 1989), $p56^{L32}$ does not require ATP for optimal binding to RNA (unpublished results). Despite their similarities, these three pPyBPs seem to be

distinct molecules. U2AF and pPTB are both nuclear proteins, whereas $p56^{L32}$ is cytosolic, with about half of its cellular activity associated with the ribosomal fraction. In unpublished experiments, using purified $p56^{L32}$ and pPTB (from A. Gil and P. Sharp), the electrophoretic mobilities of the native complexes with the same RNA probe seem to be different, as do the mobilities of the UV-cross-linked proteins under denaturing conditions. Additionally, antiserum raised against pPTB (from A. Gil and P. Sharp) did not cross-react with $p56^{L32}$. Therefore, these three activities, with very similar RNA-binding specificities, seem to reside in different protein molecules, although definitive proof of this awaits cloning of $p56^{L32}$.

Internal Translational Initiation on Viral RNA Molecules

The genomes of the picornavirus family are encoded in single-stranded RNA molecules of positive polarity. The genomic RNA of poliovirus, a well-studied member of this family, is approximately 7.5 kilobases in size and contains a single large open reading frame, which encodes the viral polyprotein. The open reading frame is preceded by a complex leader of approximately 740 nucleotides. Upon infection, poliovirus subjugates the translational apparatus of the host cell through proteolytic cleavage of the p220 component of initiation factor eIF-4F (Sonenberg, 1988; Rhoads, 1988). This initiation factor is responsible in uninfected cells for recognition of the 5′ cap structure of mRNAs; inactivation of this factor after virus infection results in blockade of cap-dependent initiation on host mRNAs. The synthesis of picornavirus proteins is resistant to destruction of eIF-4F. This resistance is due to the use of a mode of translational initiation that does not depend on cap recognition, in which ribosomes bind to the 5′ leaders of the viral RNAs at internal sites. The site of internal ribosome entry in poliovirus RNA seems to be a complex structure encompassing several hundred nucleotides [reviewed in Sonenberg (1990)], and mutational analysis has revealed that various parts of this structure interact to modulate initiation activity. In those picornavirus RNAs where the region defining internal ribosome entry has been delimited, a pPy tract has been invariably found toward the 3′ boundary. The pPy tract has been shown to be strongly required for entry of ribosomes onto the RNAs of foot-and-mouth disease virus, encephalomyocarditis virus, and poliovirus (Kuhn et al., 1990; Jang & Wimmer, 1990; Meerovitch et al., 1991; Nicholson et al., 1991).

Internal initiation on picornavirus RNAs requires factors preexisting in the host cell, and different cell types seem to vary in their levels of these factors. For example, poliovirus RNA is translated inefficiently in reticulocyte lysates, and supplementation of these reactions with extracts of HeLa or other cells promotes efficient initiation at the correct site (Brown & Ehrenfeld, 1979; Dorner et al., 1984). The factor or factors required for internal initiation in reticulocyte lysates have been partially purified from Krebs-2 cells (Svitkin et al., 1988). The existence of factors in uninfected cells that are required for internal entry of ribosomes on picornavirus messages implies that mRNAs may exist in uninfected cells that use a cap-independent mechanism of initiation. Translation of the mRNA encoding a chaparonin, glucose-regulated protein 78 (GRP78)/immunoglobulin heavy-chain binding protein (BiP), is resistant to poliovirus infection (Sarnow, 1989), and its 5′ leader has been shown to contain a site of internal ribosome entry (Macejak & Sarnow, 1991). Although the sequences in the GRP78/BiP mRNA that are necessary for internal initiation have not yet been defined, it is interesting to note that no tract of consecutive pyrimidines longer than six nucleotides in length can be identified in its 5′ leader (Ting & Lee, 1988).

Proteins have been identified in extracts from uninfected cells that cross-link to the region of internal initiation in picornavirus mRNAs (Meerovitch et al., 1989; Jang & Wimmer, 1990; Luz & Beck, 1991; Gebhard & Ehrenfeld, 1992). Meerovitch et al. (1989) reported a 52-kDa protein, present in the postribosomal supernatant of HeLa cells, that bound in a sequence-specific manner to a fragment of the 5′ leader of poliovirus RNA. Although this fragment contained the pPy tract, p52 also recognizes nucleotides adjacent to the pPy sequence (Meerovitch et al., 1989; Pestova et al., 1991; N. Sonenberg, personal communication). A second protein of 57-kDa molecular size is located in the ribosome salt-wash of a number of cell types and has been found to bind in the 5′ leader of several picornaviruses (Jang & Wimmer, 1990; Pestova et al., 1991; Luz & Beck, 1991). p57 has been shown to bind to two sites in foot-and-mouth disease virus, and one of these sites seems to be the pPy tract (Luz & Beck, 1991). However, it does not bind to the pPy tract of the encephalomyocarditis virus (Pestova et al., 1991). It is intriguing to note that p57 is similar in molecular size, subcellular localization, and RNA-binding properties to $p56^{L32}$. These may be different proteins or, alternatively, the picornaviruses may have turned $p56^{L32}$ (or another cellular pPyBP) to their own purposes in promoting internal translational initiation. If p57 is identical to one of the known cellular pPyBPs, that protein may also be involved in translation of those cellular mRNAs that use internal initiation. However, it should be reemphasized that the leader of GRP78/BiP mRNA does not contain a pPy tract longer than six nucleotides and the possible interaction of cellular pPyBPs with this RNA molecule needs to be tested.

Polyuridine and "AU-Rich" Elements

"AU-rich" sequences occur frequently in the 3′-UTRs of eukaryotic mRNAs, and the predominant nucleotide base in these regions is uridine. The predominance of polyuridine sequences is significant, since, in most of the instances discussed below, mutation of uridine residues appears to be much less tolerated than changes in adenosine residues. In the following sections we have reviewed the involvement of polyuridine elements in mRNA degradation, cytoplasmic polyadenylation, and transcriptional termination, restricting our discussion to those examples where binding proteins have been identified. By definition, polyuridine tracts are pPy elements; however, these elements lack cytosine residues, and substitution of cytosine for uridine is inhibitory in at least one case (Gillis & Malter, 1991). Although it is too early to draw definitive conclusions, these observations suggest that proteins interacting with these polyuridine tracts could have different binding requirements than those acting in conjunction with the pPy elements described above.

RNA Degradation. The stability of individual cellular mRNA species is highly variable with some mRNAs having half-lives of hours to days and others turning over in a matter of minutes [see Atwater et al. (1990) for a recent review]. The features of mRNA structure that determine constitutive and regulated degradation rate are beginning to be understood. Shaw and Kamen (1986) first demonstrated that the introduction of a sequence from the 3′-UTR of the GM-CSF mRNA into the β-globin 3′-UTR was sufficient to confer instability on the otherwise stable β-globin mRNA. The critical region of the GM-CSF 3′-UTR is uridine-rich and

contains the sequence AUUUA. This so-called AU-rich element (ARE) is present in the 3′-UTR of a number of cytokine and oncogene mRNAs (Shaw & Kamen, 1986; Caput et al., 1986) and is largely responsible for their rapid intracellular turnover. An early step in mRNA degradation seems to be assembly of a >20S complex that is dependent on the presence of the ARE (Savant-Bhonsale & Cleveland, 1992). Most AREs contain multiple copies of the AUUUA sequence (Shaw & Kamen, 1986; Caput et al., 1986), and a single AUUUA sequence does not appear to be sufficient to impart instability, unless it is flanked by polyuridine regions (Vakalopoulou et al., 1991). The importance of the uridine content of AREs is further illustrated by the observation that replacement of uridine residues with adenosine eliminated the destabilizing influence of these elements (Shyu et al., 1991).

Recently, several groups have identified proteins that interact specifically with AREs (Malter, 1989; Gillis & Malter, 1991; Vakalopoulou et al., 1991; Bohjanen et al., 1991; You et al., 1992). Malter (1989) observed that a complex formed between cytoplasmic extracts from lymphocytes and a labeled RNA containing four copies of the AUUUA sequence. The protein factor (or factors) was found to complex with AREs from GM-CSF, IL-3, interferon, c-*fos*, and c-*myc* (Gillis & Malter, 1991) but would not bind RNA containing a single AUUUA element or the sequence AUGUA reiterated four times. An extended polyuridine tract flanked by adenosine residues (AUUUUUUUA) showed high binding activity (Gillis & Malter, 1991); however, poly(U) was reported not to compete for binding (Malter, 1989; Gillis & Malter, 1991). Similar proteins have been identified in Hela and erythroleukemia cells, which bind the GM-CSF and c-*myc* AREs, respectively, although they also bind poly(U) (Brewer, 1991; Vakalopoulou et al., 1991). Despite conflicting results with poly(U), these data are generally consistent with the observations that the polyuridine content of AREs is critical for proper activity in vivo.

Although the stability of individual ARE-containing mRNAs can be regulated discoordinately in vivo (Lindsten et al., 1989; Schuler & Cole, 1988), the ARE-binding proteins described above appear to bind all AREs with roughly equal affinity. Recently, Bohjanen et al. (1991) reported a new RNA-binding protein present only in extracts of mitogen-activated lymphocytes. This protein specifically bound to the AREs of GM-CSF, IL-2, and TNF-α mRNAs but not to the AU-rich sequences contained in the 3′-UTR of c-*myc*. These same workers found a constitutive ARE-binding activity in both unstimulated and activated lymphocytes, which was different from the inducible factor. This constitutive factor appeared to be similar to the ARE-binding proteins described previously (Malter, 1989; Gillis & Malter, 1991; Vakalopoulou et al., 1991). Thus, T-lymphocytes seem to contain at least two distinct ARE-binding proteins: a constitutive factor that may participate in general destabilization of ARE-containing mRNAs and an inducible factor that is thought to specifically stabilize cytokine mRNAs in activated cells.

Cytoplasmic Polyadenylation. Polyadenylation of mRNA molecules occurs in the cytoplasm by a mechanism different from the constitutive reaction that occurs during nuclear pre-mRNA processing (Swimmer & Shenk, 1985). In immature animal oocytes, newly transcribed mRNAs are polyadenylated by the nuclear pathway; however, a fraction of these mRNAs is then rapidly deadenylated after transport to the cytoplasm (Huarte et al. 1992). These deadenylated mRNAs are unable to enter polysomes and appear to represent stored maternal mRNA species (Rosenthal et al., 1983). During specific stages of development [for a review, see Davidson (1986)], stored mRNA molecules are polyadenylated, triggering initiation of translation and entry into polysomes. The processes of cytoplasmic deadenylation (Huarte et al., 1992) and polyadenylation (Fox et al., 1989; Paris & Richter, 1990; McGrew & Richter, 1990) both appear to require the same polyuridine element, which is found at variable positions within the 3′-UTR (McGrew and Richter, 1990). Cytoplasmic polyadenylation additionally requires the conserved AAUAAA sequence, which is also necessary for nuclear polyadenylation. A comparison of mammalian and *Xenopus* cytoplasmic polyadenylation elements (CPEs) has revealed a highly conserved polyuridine element with a consensus sequence of UUUUUUAU [see Bachvarova (1992) for a recent review]. Studies have demonstrated that substitution of just two uridine residues in the CPE of *Xenopus* G10 mRNA strongly inhibits, if not abolishes, cytoplasmic polyadenylation in mature oocytes, while mutation of the conserved AU dinucleotide is only partially inhibitory (McGrew et al., 1989; McGrew & Richter, 1990). These data indicate that uridine content is a key feature of the conserved CPE sequence (McGrew & Richter, 1990).

Binding activities have been identified that interact with the CPEs of two different translationally regulated mRNAs from *Xenopus*, G10 (McGrew et al., 1989) and B4 (Paris & Richter, 1990). An 82-kDa protein binds to the G10 CPE and has been identified in extracts from mature oocytes but not from immature oocytes (McGrew & Richter, 1990). This protein will not bind to RNAs containing a mutant CPE or just the AAUAAA sequence. The CPE of mRNA B4 interacts with a 58-kDa protein. This protein was present in extracts from immature oocytes and seemed to be modified in mature oocytes, perhaps by phosphorylation controlled by cyclin (Paris et al., 1991). The nature of these two proteins, their roles in cytoplasmic polyadenylation, their binding specificities, and the mechanisms of their regulation have yet to be defined.

Termination of Transcription. Polyuridine regions serve as termination signals for a variety of DNA-directed RNA polymerases. The trans-acting factors recognizing these motifs are not known in most cases, and specific consensus nucleotide sequences have been difficult to define [for recent reviews, see Platt (1986) and Kerppola and Kane (1991)]. There are, however, two clear examples of polyuridine elements serving as termination signals. Termination by eukaryotic RNA polymerase III requires a conserved polyuridine element followed by a short G/C-rich region (Bogenhagen & Brown, 1981) and results in RNA transcripts containing a uridylate tail of three to five residues (Gottlieb & Steitz, 1989a). A 50-kDa protein (Francoeur & Mathews, 1982), identified as the autoantigen La (Gottlieb & Steitz, 1989a), was found to be associated with polymerase III transcripts (Hendrick et al., 1981; Rinke & Steitz, 1982). La-antigen specifically interacts with oligouridine and was proposed to bind the transcribed polyuridine element (van Eekelen et al., 1982; Mathews & Francoeur, 1984). Depletion of La-antigen from Hela cell extracts (Gottlieb & Stietz, 1989a,b) strongly inhibited in vitro transcription by polymerase III, and RNA molecules synthesized under these conditions did not contain a full polyuridine tail (Gottlieb & Stietz, 1989a). These data suggest that once the polyuridine site is incorporated into nascent RNA, La antigen binds, destabilizing the transcription complex and releasing the newly transcribed RNA.

Transcriptional termination by vaccinia RNA polymerase is similar to that of RNA polymerase III, in that required signals are contained in the nascent RNA transcript (Rohrmann et al., 1986; Yeun & Moss, 1987; Shuman & Moss,

1988). The required RNA sequence, UUUUUNU, is located approximately 50 nucleotides from the 3′ end of the vaccinia transcripts. A virally encoded product, the heterodimeric protein termed VTF, has been shown to be required for both termination and production of the 5′ cap structure (Schuman et al., 1987), although capping is not a prerequisite for termination (Schuman & Moss, 1988; Luo et al., 1991; Hagler & Shuman, 1992). It remains to be established whether VTF, by itself, can directly bind the polyuridine termination signal or whether this element is recognized by another molecule.

It should be noted that earlier work on rho-dependent transcriptional termination in prokaryotes had indicated that termination occurred following regions of high CU content in the mRNA [reviewed in Richardson (1990)], suggesting that rho–protein was a pPyBP. Recent studies, however, have demonstrated that cytidine and not uridine residues are critical for termination promoted by rho (Rivellini et al., 1991; Zalatain & Platt, 1992). It has been suggested that rho-protein does not recognize pPy tracts but binds cytidine tracts within the context of either uridine or adenosine, but not guanosine (Rivellini et al., 1991). This hypothesis for rho recognition is presently controversial (Zalatan & Platt, 1992).

Conclusions

It seems clear that pPy tracts and pyrimidine-rich sequences are emerging as highly significant functional motifs in RNA metabolism. The existence of these motifs raises several questions, since there seem to be a number of proteins that bind to these elements in a manner that appears not to require a specific sequence. Do all pPyBPs bind to pPy tracts without sequence specificity? Are there common structures among the pPyBPs that determine their binding specificities and define a family or families of proteins? Given similar RNA-binding specificities, what directs the individual pPyBPs to their appropriate sites of action? An ancillary question is how one can establish with certainty the biological role of a particular pPyBP, given the similar binding specificity among these proteins.

Examination of proposed pPyBP binding sites have revealed neither consensus sequences, other than pPy or polyuridine tracts, nor obvious secondary structures. Does this imply that pPyBPs have loose requirements for binding, with little sequence specificity, or are there structural features other than classic stem–loops or pseudoknots that could specify pPyBP binding to RNA? In general, there are few instances where structural requirements for binding to RNA of eukaryotic proteins have been defined [for reviews, see Frankel et al. (1991) and Keene and Query (1991)]. This may reflect both an inability to measure subtle differences in binding specificity using experimental approaches such as UV cross-linking and difficulties in recognizing and predicting RNA secondary structure (Frankel et al., 1991). One of the best studied RNA-binding proteins is *Escherichia coli* transcriptional termination factor rho. Binding of rho to RNA shows many similarities to the pPyBPs, although cytidine and not uridine seems to be the preferred nucleotide in its binding sites (Richardson et al., 1990). Like eukaryotic pPyBP binding sites, binding sites for rho show no obvious sequence conservation or stem–loop structures. However, recent work has shown that not all cytidine residues within rho binding sites are equal, suggesting that specific residues, or patterns of residues, within primary or secondary structure are important for efficient interaction of rho and resulting termination (Zalatan & Platt, 1992). Thus, there is clearly a component of rho interaction that is not understood. Whether similar subtle distinctions between the binding of pPyBPs exist should be revealed upon detailed analysis of binding specificities. Such subtle differences between binding sites could explain in part how pPyBPs might discriminate between different pPy elements.

As discussed above, the pPyBPs all show very similar binding characteristics, preferring pPy tracts, with no sequence specificity that has yet been detected. The similar binding properties of the pPyBPs may be an example of convergent evolution, or, alternatively, these proteins may be the products of a family, or limited number of families, of related genes with similar RNA-binding domains. Our structural knowledge of protein–RNA interactions is rapidly expanding but at this point remains rudimentary. One of the better understood structures is that of protein A of the U1 snRNP, which has RNA-binding domains consisting of a four-stranded β-sheet, two α-helices, and a cluster of basic amino acids (Nagai et al., 1990). This structure, the RNP-CS, has now been identified by analysis of the amino acid sequence of a number of RNA-binding proteins [discussed in Bandziulis et al. (1989) and Frankel et al. (1991)], including the pPyBP, U2AF (Zamore et al., 1992). Other modes of RNA binding are now appearing as additional protein structures are analyzed. One extreme is the example of glutaminyl-tRNA synthase, where the contacts of this protein with its RNA substrate broadly extend over several domains (Rould et al., 1989). Another interesting example is an arginine-rich sequence that has been identified in a variety of RNA-binding proteins (Lazinski et al., 1989), including the HIV-1 Tat and Rev proteins (Calnan et al., 1991; Kjems et al., 1992). In this case, the arginine residues of the binding proteins have been suggested to interact with specific conformations of the sugar–phosphate backbones of their target RNA molecules (Calnan et al., 1991; Kjems et al., 1992). The interferon-induced dsRNA-activated protein kinase (Katze, 1992) possesses yet another RNA-binding domain (McCormick et al., 1992), the detailed structure of which has not yet been defined. To date, the amino acid sequences of only two of the pPyBPs are known. As noted above, U2AF contains three RNP-CS domains, which seem to be involved in RNA binding (Zamore et al., 1992). The proteins encoded by the pPTB cDNA clones that have been isolated (Gil et al., 1991; Patton et al., 1991; Brunel et al., 1991; Bothwell et al., 1991) have little similarity to U2AF and show a loosely defined pattern of hydrophobic amino acids found within several nucleic acid binding proteins, including *Drosophila elav* proteins, poly(A) binding protein, nucleolin, *Drosophila sex-lethal*, and U1 70K protein, but no homology to other known features of RNA-binding domains. Thus, at this point we have insufficient information to do more than to raise this point of possible evolutionary relationship between at least some of the pPyBPs.

The specificity of interaction of the pPyBPs presents somewhat of an enigma. These proteins participate in a spectrum of posttranscriptional activities within cells, and yet, as far as we know, their specificity of RNA binding is dictated only by interaction with pPy tracts with no sequence specificity. How is biological and biochemical specificity generated in these interactions? A clue may reside in the fact that assembly of complex structures seems to be a common element in the diverse actions of many of the pPyBPs. For example, the particles containing the untranslated r-protein mRNAs must have many components, since they sediment at 35–45S. Spliceosomal structures are equally complex, being comprised of multiple snRNPs and ancillary splicing factors. Although the detailed structures of the internal ribosome entry regions

have not been rigorously defined, the site in the poliovirus message seems to be of considerable complexity, involving several hundred nucleotides with extensive potential secondary structure (Sonenberg, 1990). From these considerations, one suspects that the pPyBPs may not act in isolation but probably make multiple intermolecular contacts in forming multicomponent molecular assemblies. In such an assembly, the pPyBPs probably interact not only with the pPy tract of the involved RNA molecules but also with other components of the complexes, including both proteins and RNA. Thus, specificity in the binding of a particular pPyBP would be dictated by the aggregate of its interactions, not only with the pPy tract but with the other components as well. A multiple domain structure has been proposed previously for RNA-binding proteins (Bandziulis et al., 1989). It has been suggested that the RS domain of U2AF is involved in other interactions in spliceosome assembly besides binding to the pPy tract (Zamore et al., 1991). It is not unusual to find multiple RNA-binding domains in the RNP-CS-type proteins (Bandziulis et al., 1989). There is evidence that p57 (Luz & Beck, 1991) and $p56^{L32}$ (unpublished) both interact with additional regions of their RNA targets besides the pPy tracts. We envision that specificity could be generated through these multiple interactions with both proteins and RNA and that the complex of a pPyBP with its pPy tract may provide a nucleation site for formation of more complex ribonucleoprotein assemblies.

The lack of unique binding specificities creates a problem in attempting to assign biological roles to pPyBPs. For example, functions for $p56^{L32}$ and pPTB in translational control and splicing have been inferred from their binding to pPy tracts and from their cytosolic and nuclear locations, respectively. However, functional assays, either in vitro or in vivo, will be required before definite assignments of biological roles can be made. Indeed, on the basis of studies in vitro of spliceosome assembly, it appears that pPTB may not be involved in this process (Jamison et al., 1992). In the case of U2AF, the ability to complement an in vitro splicing reaction facilitated identification of the function of this pPyBP. A similar demonstration of biological function is required for $p56^{L32}$, pPTB, and others of the pPyBPs.

In closing, one should note the implications of these considerations on regulation of the elaborate biochemical processes in which the pPyBPs participate. Given the multiple molecular interactions involved, formation of the multicomponent complexes discussed above need not necessarily be regulated solely by the affinity of a pPyBP for its pPy tract. It seems equally likely that formation of these complexes could be controlled at later steps of assembly, perhaps by post-translational modifications such as phosphorylation/dephosphorylation. Indeed, binding of $p56^{L32}$ to r-protein L32 mRNA seems not to be regulated by cell growth, whereas sequestration of the mRNA into mRNP particles clearly is (Kaspar et al., 1992), consistent with regulation at a step beyond initial interaction with the pPy tract.

ACKNOWLEDGMENT

We want to express our gratitude to Anna Gil and Phillip Sharp, who graciously provided us with materials to test the relationship between pPTB and $p56^{L32}$. We are grateful to Adam Geballe, Michael Katze, and Nahum Sonenberg for helpful suggestions on the manuscript and to Nahum Sonenberg also for discussing unpublished results from his laboratory.

REFERENCES

Agrawal, M. G., & Bowman, L. H. (1987) *J. Biol. Chem. 262*, 4868–4875.

Amaldi, F., Bozzoni, I., Beccari, E., & Pierandrei-Amaldi, P. (1989) *Trends Biochem. Sci. 53*, 175–178.

Atwater, J. A., Wisdom, R., & Verma, I. M. (1990) *Annu. Rev. Genet. 24*, 519–541.

Bachvarova, R. F. (1992) *Cell 69*, 895–897.

Bandziulis, R. J., Swanson, M. S., & Dreyfuss, G. (1989) *Genes Dev. 3*, 431–437.

Bindereif, A., & Green, M. R. (1987) *EMBO J. 6*, 2415–2424.

Bogenhagen, D. F., & Brown, D. D. (1981) *Cell 24*, 261–270.

Bohjanen, P. R., Petryniak, B., June, C. H., Thompson, C. B., & Lindsten, T. (1991) *Mol. Cell. Biol. 11*, 3288–3295.

Bothwell, A. L. M., Ballard, D. W., Philbrick, W. M., Lindwall, G., Maher, S. E., Bridgett, M. M., Jamison, S. F., & Garcia-Blanco, M. A. (1991) *J. Biol. Chem. 266*, 24657–24663.

Brewer, G. (1991) *Mol. Cell. Biol. 11*, 2460–2466.

Brown, B. A., & Ehrenfeld, E. (1979) *Virology 97*, 396–405.

Brunel, F., Alzari, P. M., Ferrara, P., & Zakin, M. M. (1991) *Nucleic Acids Res. 19*, 5237–5245.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. D. (1991) *Science 252*, 1167–1171.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimmer, S., & Cerami, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1670–1674.

Chitpatima, S. T., Makrides, S., Bandyopadhyay, R., & Brawerman, G. (1988) *Nucleic Acids Res. 16*, 2350.

Davidson, E. H. (1986) in *Gene Activity in Early Development*, 3rd ed., pp 46–125, Academic Press, Inc., New York.

De Philip, R. M., Rudert, W. A., & Lieberman, I. (1980) *Biochemistry 19*, 1662–1669.

Dominski, Z., & Kole, R. (1992) *Mol. Cell. Biol. 12*, 2108–2114.

Dorner, A. J., Semler, B. L., Jackson, R. J., Hanecak, R., Duprey, E., & Wimmer, E. (1984) *J. Virol. 50*, 507–514.

Fox, C. A., Sheets, M. D., & Wickens, M. P. (1989) *Genes Dev. 3*, 2151–2162.

Francoeur, A. M., & Mathews, M. B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6772–6776.

Frankel, A. D., Mattaj, I. W., & Rio, D. C. (1991) *Cell 67*, 1041–1046.

Frendeway, D., & Keller, W. (1985) *Cell 42*, 355–367.

Freyer, G. A., O'Brien, J. P., & Hurwitz, J. (1989) *J. Biol. Chem. 264*, 14631–14637.

Fu, X.-Y., Ge, H., & Manley, J. L. (1988) *EMBO J. 7*, 809–817.

Garcia-Blanco, M. A., Jamison, S. F., & Sharp, P. A. (1989) *Genes Dev. 3*, 1874–1886.

Gebhard, J. R., & Ehrenfeld, E. (1992) *J. Virol. 66*, 3101–3109.

Geyer, P. K., Meyuhas, O., Perry, R. P., & Johnson, L. F. (1982) *Mol. Cell. Biol. 2*, 685–693.

Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C., & Dreyfus, G. (1992) *Nucleic Acids Res. 20*, 3671–3678.

Gil, A., Sharp, P. A., Jamison, S. F., & Garcia-Blanco, M. A. (1991) *Genes Dev. 5*, 1224–1236.

Gillis, P., & Malter, J. S. (1991) *J. Biol. Chem. 265*, 3172–3177.

Gottlieb, E., & Steitz, J. A. (1989a) *EMBO J. 8*, 841–850.

Gottlieb, E., & Steitz, J. A. (1989b) *EMBO J. 8*, 851–861.

Grabowski, P. J., & Sharp, P. A. (1986) *Science 233*, 1294–1299.

Hagler, J., & Shuman, S. (1992) *Science 255*, 983–986.

Hammond, M. L., Merrick, W., & Bowman, L. H. (1991) *Genes Dev. 5*, 1723–1736.

Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R., & Steitz, J. A. (1981) *Mol. Cell. Biol. 1*, 1138–1149.

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., & Shimura, Y. (1991) *Science 252*, 833–836.

Huarte, J., Stutz, A., O'Connel, M. L., Gubler, P., Belin, D., Darrow, A. L., Strickland, S., & Vassalli, J.-D. (1992) *Cell 69*, 1021–1030.

Inoue, K., Hoshijima, K., Sakamoto, H., & Shimura, Y. (1990) *Nature 344*, 461–463.

Jamison, S. F., Crow, A., & Garcia-Blanco, M. A. (1992) *Mol. Cell. Biol. 12*, 4279–4287.

Jang, S. K., & Wimmer, E. (1990) *Genes Dev. 4*, 1560–1572.

Jansen-Durr, P., Boshart, M., Lupp, B., Bosserhoff, A., Frank, R. W., & Schutz, G. (1992) *Nucleic Acids Res. 20*, 1243–1249.

Kaspar, R. L., Rychlik, W., White, M. W., Rhoads, R. E., & Morris, D. R. (1990) *J. Biol. Chem. 265*, 3619–3622.

Kaspar, R. L., Kakegawa, T., Cranston, H., Morris, D. R., & White, M. W. (1992) *J. Biol. Chem. 267*, 508–514.

Kaspar, R. L., Morris, D. R., & White, M. W. (1993) in *Translational Regulation of Gene Expression II* (Ilan, J., Ed.) Plenum Publishing Corp., New York.

Katze, M. G. (1992) *J. Interferon Res. 12*, 241–248.

Keene, J. D., & Query, C. C. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 179–202.

Kerfelec, B., LaForge, K. S., Vasiloudes, P., Puigserver, A., & Scheele, G. A. (1990) *Eur. J. Biochem. 190*, 299–304.

Kerppola, T. K., & Kane, C. M. (1991) *FASEB J. 5*, 2833–2842.

Kjems, J., Calnan, B. J., Frankel, A. D., & Sharp, P. A. (1992) *EMBO J. 11*, 1119–1129.

Krainer, A. R., & Maniatis, T. (1988) in *Molecular Biology: Transcription and Splicing* (Hames, B. D., & Glover, D. M., Eds.) pp 131–206, IRL Press, Oxford.

Kuhn, R., Luz, N., & Beck, E. (1990) *J. Virol. 64*, 4625–4631.

Levy, S., Avni, D., Hariharan, N., Perry, R. P., Meyuhas, O. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3319–3323.

Lindsten, T., June, C. H., Ledbetter, J. A., Stella, G., & Thompson, C. B. (1989) *Science 244*, 339–343.

Luo, Y., Hagler, J., Shuman, S. (1991) *J. Biol. Chem. 266*, 13303–13309.

Luz, N., & Beck, E. (1991) *J. Virol. 65*, 6486–6494.

Macejak, D. G., & Sarnow, P. (1991) *Nature 353*, 90–94.

Makrides, S., Chitpatima, S. T., Bandyopadhyay, R., & Brawerman, G. (1988) *Nucleic Acids Res. 16*, 2349.

Malter, J. S. (1989) *Science 246*, 664–666.

Mariottini, P., & Amaldi, F. (1990) *Mol. Cell. Biol. 10*, 816–822.

Mathews, M. B., & Francoeur, A. M. (1984) *Mol. Cell. Biol. 4*, 1134–1140.

McCormick, S. J., Thomis, D. C., & Samuel, C. E. (1992) *Virology 188*, 47–56.

McGrew, L. L., & Richter, J. D. (1990) *EMBO J. 9*, 3743–3751.

McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., & Richter, J. D. (1989) *Genes Dev. 3*, 803–815.

Meerovitch, K., Pelletier, J., & Sonenberg, N. (1989) *Genes Dev. 3*, 1026–1034.

Meerovitch, K., Nicholson, R., & Sonenberg, N. (1991) *J. Virol. 65*, 5895–5901.

Meyuhas, O., Thompson, E. A., & Perry, R. P. (1987) *Mol. Cell. Biol. 7*, 2691–2699.

Mullen, M. P., Smith, C. W. J., Patton, J. G., & Nadal-Ginard, B. (1991) *Genes Dev.* 642–655.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990) *Nature 348*, 515–520.

Nicholson, R., Pelletier, J., Le, S.-Y., & Sonenberg, N. (1991) *J. Virol. 65*, 5886–5894.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. R., & Sharp, P. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.

Paris, J., & Richter, J. D. (1990) *Mol. Cell. Biol. 10*, 5634–5645.

Paris, J., Swenson, K., Piwnica-Worms, H., & Richter, J. D. (1991) *Genes Dev. 5*, 1697–1708.

Patterson, B., & Guthrie, C. (1991) *Cell 64*, 181–187.

Patton, J. G., Mayer, S. A., Tempst, P., & Nadal-Ginard, B. (1991) *Genes Dev. 5*, 1237–1251.

Pestova, T. V., Hellen, C. U. T., & Wimmer, E. (1991) *J. Virol. 65*, 6194–6204.

Pinsky, S. D., LaForge, K. S., & Scheele, G. (1985) *Mol. Cell. Biol. 5*, 2669–2676.

Platt, T. (1986) *Annu. Rev. Biochem. 55*, 339–372.

Reed, R. (1989) *Genes Dev. 3*, 2113–2123.

Reed, R., & Maniatis, T. (1985) *Cell 41*, 95–104.

Rhoads, R. E. (1988) *Trends Biochem. Sci. 13*, 52–56.

Richardson, J. P. (1990) *Biochim. Biophys. Acta 1048*, 127–138.

Rinke, J., & Steitz, J. A. (1982) *Cell 29*, 149–159.

Rivellini, F., Alifano, P., Piscitelli, C., Blasi, V., Bruni, C. B., & Carlomagno, M. S. (1991) *Mol. Microbiol. 5*, 3049–3054.

Rohrmann, G., Yuen, L., & Moss, B. (1986) *Cell 46*, 1029–1035.

Rosenthal, E. T., Tansey, T., & Ruderman, J. R. (1983) *J. Mol. Biol. 166*, 309–327.

Rould, M. A., Perona, J. J., Soll, D., & Steitz, T. A. (1989) *Science 246*, 1135–1142.

Ruskin, B., & Green, M. R. (1985) *Cell 43*, 131–142.

Sarnow, P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5795–5799.

Savant-Bhonsale, S., & Cleveland, D. W. (1992) *Genes Dev. 6*, 1927–1939.

Schuler, G. D., & Cole, M. D. (1988) *Cell 55*, 1115–1122.

Schuman, S., & Moss, B. (1988) *J. Biol. Chem. 263*, 6220–6225.

Schuman, S., Broyles, S., & Moss, B. (1987) *J. Biol. Chem. 262*, 12372–12380.

Shaw, G., & Kamen, R. (1986) *Cell 46*, 659–667.

Sharp, P. (1987) *Science 235*, 766–771.

Shyu, A., Belasco, J. G., & Greenberg, M. E. (1991) *Genes Dev. 5*, 221–231.

Smith, C. W. J., Patton, J. G., & Nadal-Ginard, B. (1989) *Annu. Rev. Genet. 23*, 527–577.

Sonenberg, N. (1988) *Prog. Nucleic Acid Res. Mol. Biol. 35*, 173–207.

Sonenberg, N. (1990) *Curr. Top. Microbiol. Immunol. 161*, 23–47.

Svitkin, Y. V., Pestova, S. V., Maslova, S. V., & Agol, V. I. (1988) *Virology 166*, 394–404.

Swimmer, C., & Shenk, T. (1985) *Nucleic Acids Res. 13*, 8053–8063.

Tian, M., & Maniatis, T. (1992) *Science 256*, 237–240.

Ting, J., & Lee, A. S. (1988) *DNA 7*, 275–286.

Tushinski, R. J., & Warner, J. R. (1982) *J. Cell. Physiol. 112*, 128–135.

You, Y., Chen, C.-Y. A., & Shyu, A. B. (1992) *Mol. Cell. Biol. 12*, 2931–2940.

Vakalopoulou, E., Schaack, J., & Shenk, T. (1991) *Mol. Cell. Biol. 11*, 3355–3364.

van Eekelen, C., Buijtels, H., Linne, T., Ohlsson, R., Philipson, L., & van Venrooij, W. (1982) *Nucleic Acids Res. 10*, 3039–3052.

Wagner, M., & Perry, R. P. (1985) *Mol. Cell. Biol. 5*, 3560–3576.

Wang, J., & Pederson, T. (1990) *Nucleic Acids Res. 18*, 5995–6001.

Yuen, L., & Moss, B. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6417–6421.

Zalatan, F., & Platt, T. (1992) *J. Biol. Chem. 267*, 19082–19088.

Zamore, P. D., & Green, M. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9243–9247.

Zamore, P. D., & Green, M. R. (1991) *EMBO J. 10*, 207–214.

Zamore, P. D., Patton, J. G., & Green, M. R. (1992) *Nature 355*, 609–614.