RNA-binding proteins as regulators of gene expression

Haruhiko Siomi* and Gideon Dreyfuss†

A plethora of post-transcriptional mechanisms are involved in essential steps in the pathway of genetic information expression in eukaryotes. These processes are specified by *cis*-acting signals on RNAs and are mediated by specific *trans*-acting factors, including RNA-binding proteins and small complementary RNAs. Recent information has begun to define the molecular mechanisms by which RNA-binding proteins recognize specific RNA sequences and influence the processing and function of RNA molecules.

Addresses
Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148, USA
*e-mail: hsiomi@hmivax.humgen.upenn.edu
†e-mail: gdreyfuss@hmivax.humgen.upenn.edu

Current Opinion in Genetics & Development 1997, 7:345–353

http://biomednet.com/elecref/0959437X00700345

Abbreviations

3′-UTR	3′-untranslated region
dsRBD	double-stranded RNA-binding domain
KH	K homology
Prm1	protamine 1
RBD	RNA-binding domain
RNP	ribonucleoprotein
Sam68	Src-associated in mitosis 68kDa

Introduction

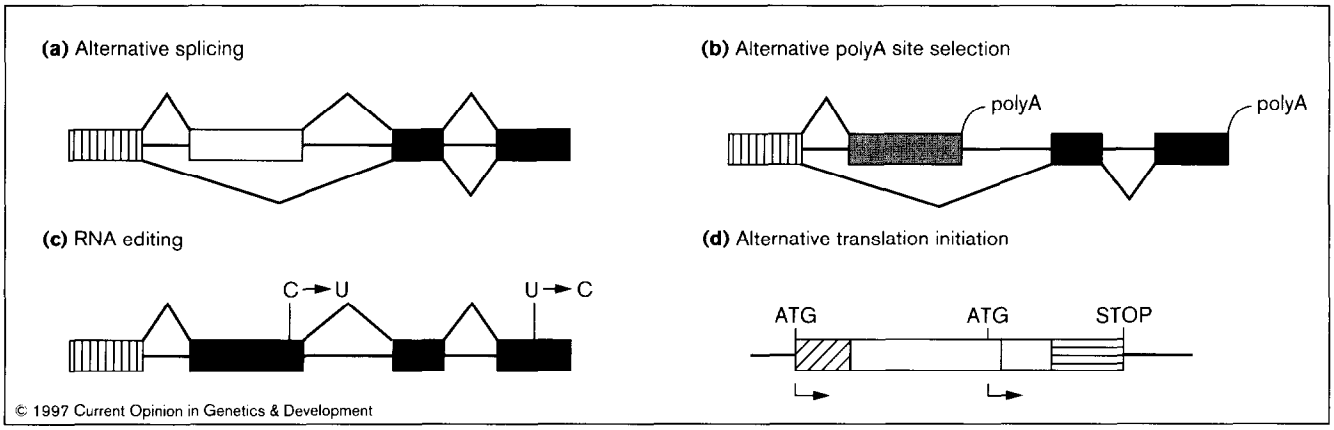
Eukaryotic cells possess an impressive repertoire of tools for the regulation of gene expression. For decades, influenced largely by the precedent of the prokaryotic paradigm, the thinking of how eukaryotic cells control the activity of their genes has focused on the selection and activation of transcriptional promoters. The view that transcriptional regulation is the predominant regulatory mechanism, however, has been increasingly challenged by the discovery of ever-increasing examples of post-transcriptional mechanisms for regulating gene expression. Post-transcriptional regulation of gene expression can involve the on/off regulation of particular gene products in a temporally and spatially regulated manner, allowing cells of different types or at different developmental stages to fine-tune their patterns of gene expression. On/off regulation of RNA can allow a cell to respond to environmental cues more quickly than *de novo* transcription permits. In fact, many important events in development — such as pattern formation and terminal differentiation — are regulated by an array of post-transcriptional mechanisms, controlling mRNA stability, localization, and translation [1–3]. Post-transcriptional regulation of gene expression can also generate an enormous range of protein products from a single gene (Fig. 1). Three forms of intranuclear fine-tuning to generate RNA sequence diversity are well established: alternative splicing of pre-mRNAs [4], alternative polyadenylation site selection [5], and RNA editing [6]. In the cytoplasm, utilization of alternative translational start sites can also produce functionally different proteins from a single mRNA [7]. In some cases, these processes constitute a switch that dramatically reverses function; (e.g. transcriptional activators and repressors can be generated from the same gene [8]). Using these sophisticated strategies, vertebrate cells can produce a much larger variety of proteins than the number of genes in the genome. Aberrant RNA processing could result in the synthesis of deleterious proteins which cause disease as a result of loss of function or by a dominant negative mechanism, or to make an otherwise useful protein at the wrong time.

RNA-binding proteins play central roles in the post-transcriptional regulation of gene expression. These proteins contain regions which function as RNA-binding domains, and auxiliary domains that mediate protein–protein interaction and subcellular targeting [9,10]. In this review, we summarize current information on the structure and function of the major types of RNA-binding domains and discuss how their activities are integrated into the general scheme of post-transcriptional gene expression.

RNA-binding domains

Work published within the past several years has significantly extended our understanding of the structure and function of the RNA-binding motifs (Table 1). One of the best-studied RNA-binding domains is the RNP motif, also referred to as RBD or RRM [9–11]. At present, there are nearly 300 known RNP motif proteins. Each RNP motif contains two consensus sequences, RNP1 and RNP2, that are part of the RNA-binding surface of this domain. All RNP motif structures determined to date reveal a large central antiparallel β sheet which is packed against two α helices, and follow a general pattern of β–α–β–β–α–β (Fig. 2a). The conserved RNP1 and RNP2 sequences are juxtaposed on the two central β strands with the side chains of conserved aromatic amino acids of RNP1 and RNP2 displayed on the surface of the β sheet. The crystal structure of the RNP domain of the U1 A protein in complex with its cognate U1 RNA hairpin has provided the most detailed structural basis for the RNA interaction of this motif [12]. The antiparallel β sheet forms a non-sequence specific RNA-binding platform made up of general stacking interactions with the bases, where the RNA is situated in a shallow RNA-binding platform rather than sequestered into a deep cleft. The loop between the second and third β strands (Loop 3) is rich in basic residues and makes up an electropositive edge to the platform, which plays a critical role in specific binding by protruding

Figure 1



Generation of protein products with clearly separable functions from a single gene. The primary transcripts of eukaryotic RNA polymerase II, termed hnRNAs or pre-mRNAs, undergo a complex and highly regulated series of events in the nucleus as they mature into functional, cytoplasmic mRNAs [9]. Recent experiments suggest that an mRNA 'factory' exists which carries out coupled transcription, splicing and cleavage-polyadenylation of mRNA precursors [55•]. This complex is effectively an assembly line that enhances mRNA production by channelling precursors directly from the synthetic machinery to the processing machinery. Exons are represented by boxes with different shading and hatching, introns by lines. Although exons do not necessarily correspond to functional domains on proteins, for convenience and simplicity, exons in this figure correspond to functional domains such as for nucleic acid binding, protein–protein interaction, subcellular targeting, enzyme catalytic sites and so on. Alternative utilization of these functional domains would produce proteins with diverse function from a single gene. For example, transcriptional activators and repressors can be encoded by the same gene in a way that a protein possessing a DNA-binding domain but lacking an activation domain can compete with activators for binding to the same site and thereby block activation [8]. **(a)** Generation of proteins with diverse function from a single gene by alternative splicing. Inclusion or exclusion of the functional domain exons during splicing generates mRNAs encoding functionally different proteins: activators or repressors, proteins with different cellular localizations, or proteins with different interaction partners. **(b)** Generation of alternative proteins by selecting different poly(A) sites. Splicing patterns and poly(A) signals are indicated. Activating and inhibiting effects of splice signals or changes in cleavage stimulation factor concentration modulate poly(A) site selection [5,56••]. Poly(A) site regulation produces not only alternative protein products but also mRNAs with different 3′-UTRs, which may alter the stability, translation, or localization of the mRNA, although the same protein will be produced. **(c)** Generation of functionally different proteins by RNA editing. RNA editing is a process in which select nucleotide sequences in RNA are altered from that originally encoded in the genome and can be divided into insertion or deletion editing, and substitution or modification editing [6]. RNA editing of pol II transcripts occurs in the nucleus. Interestingly, however, the precise site of editing of each gene may differ within the nucleus. For example, glutamate receptor subunit B pre-mRNA is edited from a glutamine (CAG) to an arginine (CGG) codon. This so-called Q/R site editing occurs before splicing on double-stranded RNA forms complementary intron and exon sequences. The tissue-specific editing of apolipoprotein B mRNA converts a glutamine (CAA) to a stop (UAA) codon, which occurs subsequent to RNA splicing. As nonsense mutations have been associated with the skipping of specific constitutively spliced exons in some genes [57], the latter event (CAA→UAA) occurs after splicing probably because of avoidance of the so-called nonsense codon mediated exon-skipping process [57]. **(d)** Generation of functionally different proteins by alternative translational start sites. There is a way for two or more proteins to be translated from a single mRNA; initiation at first and second AUGs generates long and short (lacking the upstream functional domains) protein isoforms from the same reading frame, and initiation at first and second AUGs in different, but overlapping, reading frames produces two unrelated proteins [7].

into the RNA loop (which can otherwise adopt multiple conformations), preventing the pairing of bases within the RNA and making them available to interact, in a specific fashion, with residues from the β strands of the RNP domain. Upon complex formation, the flexible RNA loop becomes ordered. An 'induced-fit' mechanism is thus the best explanation for the observed structural adaptation.

It appears that many RNA-binding proteins use β sheets as surfaces for RNA binding and that stacking interactions between nucleotide bases and aromatic side chains on the β sheets play an important role in stabilizing the complex. Protein-induced conformational changes of RNA, which involve unstacking of bases to achieve specific recognition, also seem to be a common theme seen in many RNA-binding proteins. Folding patterns similar to the RNP domain have been observed in other RNA-binding proteins, including several ribosomal proteins which lack any sequence similarity to RNP motifs, suggesting a possible common RNA-binding protein as the ancestor for both the RNP motif family and several of the ribosomal proteins [13]. The recently solved S1 domain (for ribosomal protein S1) is comprised of a five-stranded antiparallel β barrel, and aromatic amino-acid side chains are also found on the surface of the β sheet, suggesting that this domain binds RNA in a manner similar to that of the RNP domain [14].

As both the major and the minor grooves of RNA are too narrow to form a smooth surface for interactions with α helices (a situation which is dramatically different from the smooth contact between α helices and the major groove of DNA), it was suggested that the use of a β sheet in protein–RNA complexes is a widely recurring theme. This theme is clearly not a universal solution, however, as recent findings show that helices also have important

Table 1

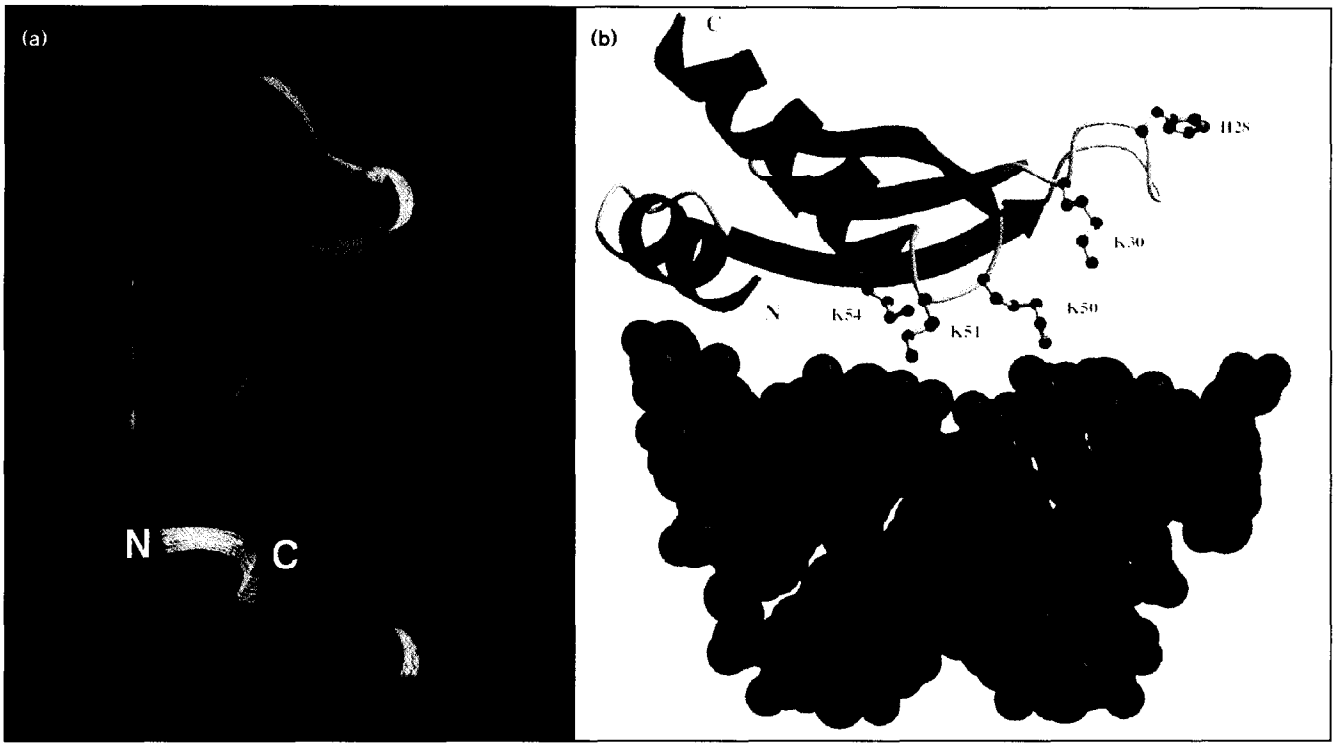
Structure of RNA-binding domains.

Motif	Topology/ secondary structure	Comments	References
RNP	β–α–β–β–α–β	The four-stranded antiparallel β sheet is packed against the two helices; the β sheet makes a flat, solvent-exposed RNA-binding surface and the second and third β sheet makes up an electropositive edge to the platform	[12,58]
KH	β–α–α–β–β–α	The three-stranded antiparallel β sheet is packed against the three helices; crosslinking and sequence conservation suggest a potential surface for RNA binding centered on the loop between the first two helices	[18•]
dsRBD	α–β–β–β–α	The three-stranded antiparallel β sheet is packed against the two helices; mutagenesis suggests that dsRNA binds in the cleft between the amino terminus of the second helix and one face of the β sheet	[15•,16•]
S1	β–β–β–α–β–β	The structure of this domain, a five-stranded antiparallel β barrel, is very similar to that of the cold-shock protein	[14•]
Zinc finger (classic)	β–β–α	The two-stranded antiparallel β sheet is packed against the helix; although the α-helical element lying in the major groove of the DNA is the primary determinant of DNA binding, little is known about the RNA-binding surface	[62]
RGG box	β-spiral	Unstacks RNA bases and acts in destabilizing RNA secondary structures	[63]
Arginine-rich motif*			
BIV Tat	β-hairpin	The Tat peptide in a β-sheet conformation penetrates deep within a widened major groove of Tar RNA in an edgewise orientation	[64]
HIV Rev	α-helical	The Rev peptide in an α-helix conformation penetrates deep into a a widened major groove of RRE	
Ribosomal S6	β–α–β–β–α–β	The four-stranded antiparallel β sheet is packed against the two helices; the folding scheme is identical to the topology of the RNP-type RBD	[66]
Ribosomal S15	α–α–α–α–α	The surface exposed by loop II and the carboxy-terminal helices III and IV is suggested to interact with rRNA	[19]
Ribosomal L11	α–β–α–α–β	Three α helices with two short-stranded parallel β sheet: the general arrangement of the three helices is suggestive of the homeodomain DNA-binding motif	[20,21•]
Homeodomain (bicoid)	α–α–α	The end of the third helix of the bicoid homeodomain contains the basic, arginine-rich peptide that may provide a Rev-like α-helical scaffold	[22••]
Rop (Rom)	α–α	Homodimer with a two-fold axis of symmetry forms the four-helix bundle; the RNA-binding determinants form a relatively flat recognition surface	[67]
APOBEC1 (p27)	α–β–α	The zinc-containing cytidine deaminase binds AU-rich sequences; zinc-binding is not necessary for RNA-binding	[68]

For more detailed information about RNA–protein interactions, see [9–11]. A number of enzymes interact with RNA. Hentze [69] identified a striking common denominator in that their catalytic reactions involve mono- or dinucleotides as substrates or cofactors, or that their structures contain occult nucleotide-binding sites such as those found in aconitase/iron-regulatory protein-1 and catalase. It has been suggested that the relationship between (di)nucleotide-binding and RNA-binding provides an evolutionary pathway for the development of RNA-binding from a more simple mono- or dinucleotide-binding function. *The arginine-rich motif may be a particularly versatile framework for recognizing RNA structures as the λ phage N peptide binds to its RNA site in an α-helical conformation and the HIV Tat peptide may bind in an extended conformation [53•].

roles in RNA-binding. The following examples serve to illustrate this point. The double-stranded RNA-binding domain (dsRBD) is a 65 amino acid motif that is found in a variety of proteins that interact with double-stranded RNAs. Structures for two dsRBDs, the *Drosophila* staufen [15•] and *Escherichia coli* RNase III [16•], have been reported. Although the structure with RNA ligands has not been determined, the basic fold consists of three antiparallel β strands that pack to form two α helices on one face (Fig. 2b), which bears some resemblance to the RNP domain. Mutational analysis, however, suggests that the well-conserved basic residues at the amino terminus of the second α helix may have a direct role in RNA-binding and that dsRNA interacts with one face of the domain, a cleft between the amino terminus of the second helix and one face of the second β sheet.

Figure 2



Structures of two commonly found RNA-binding domains. **(a)** The RNP domain of the hnRNP C protein. The RNP domain folds into an α–β–α–β–β–α–β structure, forming an antiparallel four-stranded β sheet that is packed against two perpendicularly oriented α-helices which are positioned on one side of the β sheet. The RNP1 and RNP2 consensus sequences are juxtaposed on the two central β strands (β3 and β1) of the folded domain. The residues colored red probably come directly into contact with the RNA. The binding specificity of RNP domains is found to reside mainly in the variable regions of the loops connecting the β-strands and in the terminal portions [13,58]. (Adapted from [58] and redrawn by Matthias Görlach.) **(b)** The dsRBD of *Drosophila* staufen. The dsRBD has a very similar topology to the amino-terminal domain of ribosomal protein S5 [15•]. The two domains also have similarity at the sequence level. The minimal segments of dsRNA needed for binding is 11 base pairs (bp), equivalent to one turn of A-form dsRNA. Mutation experiments suggest that RNA interacts with one face of the domain. The five amino acids that are likely to interact with RNA directly are indicated. The dsRNA shown in the lower half of the figure is a space filling representation of a 14 bp A-type helix. Hydroxyl-radical footprinting experiments using the dsRBD of the human dsRNA-activated protein kinase suggest that recognition of dsRNA by the dsRBD involves a series of minor-groove 2′-OH interactions [59]. (Adapted from [15•] and redrawn by Stefan Grünert and Mark Bycroft.)

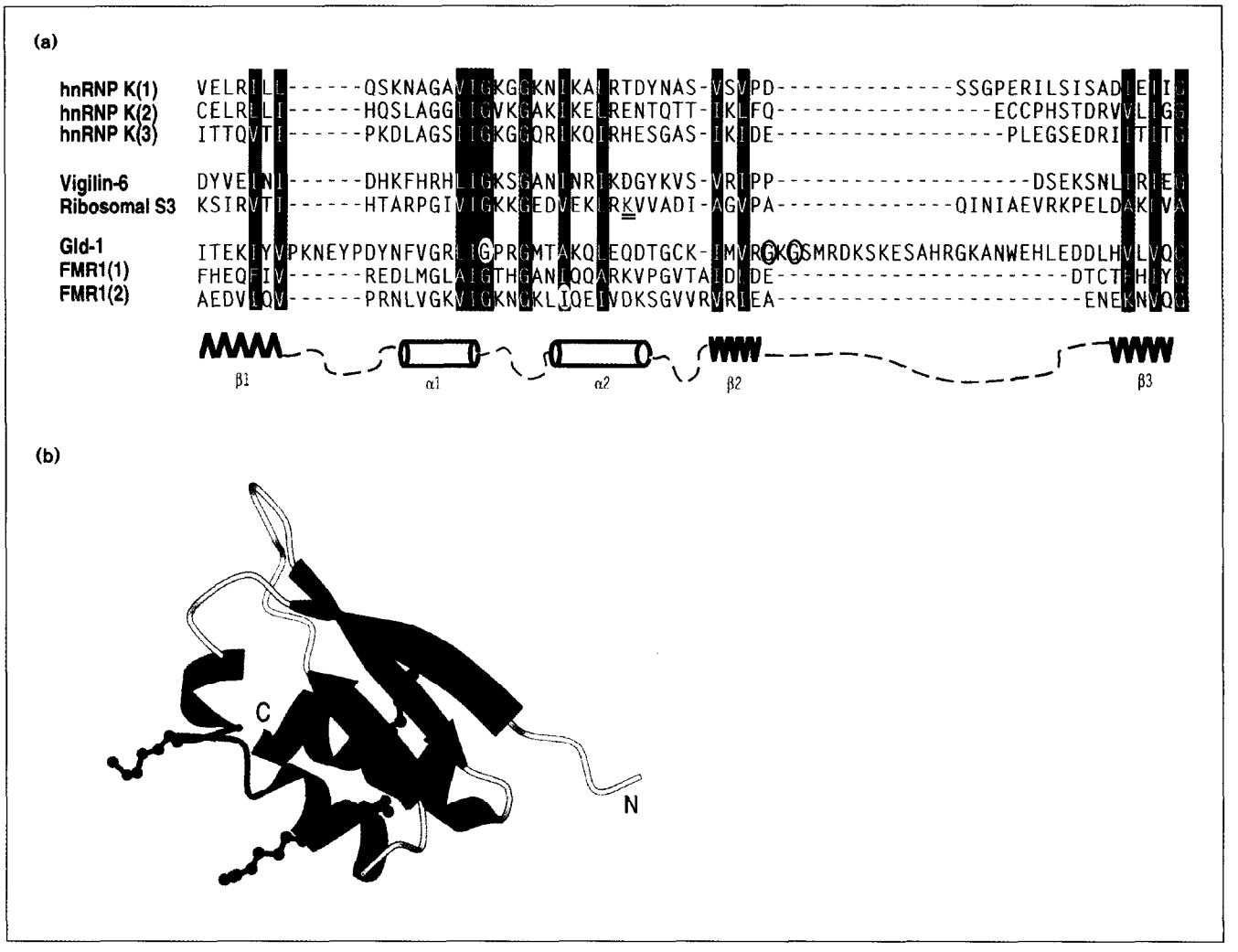
The KH domain (for hnRNP K homology domain) is comprised of ~60 amino acids and is found in a wide variety of RNA-associated proteins (Fig. 3a). The hnRNP K protein, a major pre-mRNA-binding protein, has three such motifs and was the first protein in which the motif was recognized [17]. The structure of one of the KH domains of human vigilin has been determined by NMR spectroscopy [18•]. The KH domain contains three antiparallel β sheets, which are packed against three α helices on one face of the β sheet (Fig. 3b). Sequence conservation and UV-induced RNA–protein crosslinking experiments suggest that the helical side of the KH domain interacts with RNA and the loop between the first two α helices — which contains a strongly conserved tetrapeptide, Gly–X–X–Gly, where X varies for different KH domains but is often a positively charged amino acid — plays an important role in RNA binding. The loop may penetrate the widened groove or loop of RNA and therefore, because of steric hindrance, a side chain could not be accommodated at the positions of the glycines.

Very recently, the solution structures of two ribosomal proteins, S15 and L11, have been determined. S15 is a protein that binds rRNA directly and is solely α helical in structure [19]. The rRNA-binding domain of L11 contains three α helices, the arrangement of which is strikingly similar to the homeodomain DNA-binding motif [20•,21•]. In this context, it is of particular interest that the *Drosophila* homeodomain protein bicoid represses translation of another homeodomain protein, caudal, by directly binding the 3′-untranslated region (3′-UTR) of the caudal mRNA [22••]. As bulges and mismatches within an RNA helix can considerably alter groove dimensions, these recent data suggest that an irregular RNA helix could provide a contact surface for an α helix.

Function

How does the binding of a protein to RNA regulate gene expression? Binding may alter RNA structure to facilitate or hinder interactions with positive or negative *trans*-acting factors (proteins or complementary RNAs)

Figure 3



The hnRNP K homology (KH) motif. **(a)** Sequence alignment of the KH motif-containing proteins. Highly conserved positions are highlighted on a black background. The secondary structure indicated for the KH-6 domain of human vigilin [18•] is also shown at the bottom. Note that this alignment does not show the third α helix region. The position of a single point mutation in FMRI(2) resulting in an Ile304→Asp found in a very severe fragile X-syndrome patient [39] is circled. Sites of known mutations in Gld-1 are also circled. All missense mutations in a *C. elegans* gene *Hex-3*, which may contribute to anterior–posterior differences in muscle development, are also found in the KH domain [60]. The site in ribosomal S3 that crosslinks to rRNA [61] is double underlined. **(b)** The structure of one of the human vigilin KH domains. The structure shows a mixed αβ fold similar to those of the RNP domain and the dsRBD, although the topologies are very different from one another. The strongly conserved G–X–X–G loop between helices α1 and α2 is indicated in green. Conserved positively charged residues which are likely to interact with RNA are colored light purple. Note that RNA binding could therefore extend to other parts of the α-helical side in addition to the G–X–X–G loop. Interestingly, substitution of a highly conserved hydrophobic residue located on the second helix of the vigilin KH domain (Ile32→Asn) totally destroys the KH fold; a mutation at this position in FMRI (Ile304→Asn) is responsible for the aggravated phenotype of fragile X. (Adapted from [18•] and redrawn by Giovanna Musco.)

by bringing together disparate RNA sequences or by preventing the formation of higher-order RNA structure [23]. In addition, binding proteins may provide localization or targeting signals for transport of RNA molecules to distinct intracellular locations [24].

Signal transduction and activation of RNA

Developmental cues often regulate processing of pre-mRNAs to generate functionally distinct protein isoforms or to switch gene expression on or off. The regulation of RNA by signalling could allow a cell to respond very rapidly to external stimuli, much faster than protein expression from *de novo* transcription. For example, alternative splicing can switch a transcriptional activator to a repressor, alter cell adhesion and cell type, and determine sexual differentiation [8,25]. In addition, during many crucial steps of development — such as in the specification of cell fates and pattern formation and differentiation of red blood cells and spermatocytes — the nucleus is transcriptionally silenced. As a result, developmental regulation of gene expression at these stages is dependent on cytoplasmic events. Specific mRNAs are stored in the

cytoplasm as mRNA–protein complexes and are repressed through the action of proteins that mask them from the translational apparatus. In response to a stimulus, the masking proteins are removed or modified and the mRNA is translated [2,3].

How, then, might a signalling pathway convey messages directly to primary gene transcripts? Cellular signalling pathways could control, possibly via phosphorylation, the function of *trans*-acting factors that preside over RNA processing. The RNA-binding activity of a protein involved in such a pathway could—for example upon phosphorylation, be modified, either by altering binding affinity or changing specificity of a protein–RNA interaction. The consequences of such modulations might be to activate translation of the unmasked RNAs or to prepare a protein for interaction with specific mRNAs to help target the mRNA to a specific site in the cytoplasm. Phosphorylation and dephosphorylation may also influence RNA-binding proteins by changing their subcellular location: the transported protein may direct RNA to a special cellular compartment, or may bind to specific RNAs and change their accessibility to RNA processing machineries such as the splicesome. Recent work has identified a transcription factor, Hac1p, in the yeast *Saccharomyces cerevisiae*, as a component of a pathway that signals to the nucleus the presence of unfolded proteins in the endoplasmic reticulum and has shown that the regulation of Hac1p synthesis occurs through activation of pre-mRNA splicing which appears to be mediated by a novel splicing mechanism involving a tRNA ligase [26••,27••]. It is also noteworthy that a mitotic target of c-Src kinase is a KH-motif-containing RNA-binding protein named Sam68 (for Src-associated in mitosis 68 kDa) [28,29]. The RNA-binding activity of this protein seems to be negatively regulated by tyrosine phosphorylation [30•], suggesting that the Src-mediated signalling pathway may involve activation of RNA through modulation of its RNA-binding protein substrate [31].

The 3′-UTR as an important repository of regulatory sequences

Although many nuclear RNA processing reactions are mediated by interactions between proteins and relatively short *cis*-acting elements which can be found almost anywhere along the RNA molecule, many elements that regulate expression of mRNAs in the cytoplasm reside in the 3′-UTRs. The spatially regulated synthesis of proteins is important in cell organization, for example, in the generation of a polarized cell type and in the assembly of macromolecular structures. Although a variety of spatial patterns of proteins can be achieved by post-translational sorting of the proteins from their sites of synthesis, it has become increasingly clear that the localization of mRNAs constitutes an important means of localizing proteins [1]. Localized protein synthesis is a very efficient way not only to ensure correct protein positioning but also to prevent deleterious protein–protein interactions from occurring elsewhere in the cell. The *cis*-acting signals for mRNA localization have been identified for a number of mRNAs and all, without exception so far, lie within the 3′-UTR.

It has long been known that mRNA storage and translational control play prominent roles in many developmental steps [2,3,22••]. The precise timing of translation of the stored mRNAs which are sequestered in cytoplasmic ribonucleoprotein particles also depends on the binding of *trans*-acting factors to sequences in the 3′-UTR because alteration in the timing of protein synthesis can be achieved by modifying the 3′-UTR. For example, transgenic mice that carry a protamine 1 transgene (*Prm1*) lacking its 3′-UTR show premature translation of *Prm1* mRNA, which results in a failure to produce mature sperm [32•]. It appears that relatively large sections (>40 nucleotides) of the 3′-UTRs are required for localization and translational regulation of mRNAs, suggesting that these signals are comprised of complex structures rather than short sequence motifs. The prevalence of regulatory elements in 3′-UTRs for a myriad of developmental processes may reflect the fact that 3′-UTRs are unconstrained in evolution and so provide fertile ground for the derivation of new regulatory elements [3].

Dysfunction

The essential role of RNA–protein interactions for normal cell function is highlighted by the severity of the defects that result when these systems are perturbed (e.g. see [33]). Post-transcriptional control is especially important towards the end of spermatogenesis as the spermatid pronucleus is highly condensed and transcriptionally inactive. Indeed, chromosomal deletions of genes encoding RNP motif proteins result in azoospermia (no sperm in semen) in humans and *Drosophila* [34,35•,36]. A particularly devastating example is that of the fragile X syndrome gene *FMR1* [37] which encodes a cytoplasmic ribosome-associated RNA-binding protein [33,38]. Mutations in *FMR1* lead to fragile X syndrome, the most common form of inherited mental retardation [33,38]. Although most of these molecular alterations appear to be associated with transcriptional silencing of the *FMR1* gene [33,38], in one case that has a severe phenotype, the *FMR1* protein is expressed but carries a point mutation (Ile304→Asn) in the second KH domain (Fig. 3a) [39]; this mutation severely compromises the RNA-binding activity of FMR1 [40]. Interestingly, KH motif mutations in the *Bicaudal-C* gene in *Drosophila* and the *Gld-1* gene in *C. elegans* also lead to a phenotype that is more severe than the null phenotype [41,42]. A point mutation in the human vigilin construct equivalent to Ile304→Asn in FMR1 has recently been shown by NMR analysis in which the KH fold is destroyed totally [18•]. This result might explain why mutations within the KH domain have stronger phenotypes than those observed for loss of gene expression; namely, these mutations might not only abolish their RNA-binding activity but also expose, and thereby deregulate, other functional domains such as those

involved in protein–protein interactions. These mutant proteins could titrate out a limiting factor required for normal development, thereby possibly acting as dominant negative mutants.

Within the past several years, the molecular basis of more than 10 human genetic disorders — including fragile X syndrome, myotonic dystrophy and Huntington's disease — has been established as the expansion of a simple triplet nucleotide repeat from less than 15 copies of the repeat in normal individuals to scores of copies in affected cases [43]. Although, in the case of fragile X syndrome, the gene *FMR1* itself encodes an RNA-binding protein, recent experiments suggest that interaction between RNA transcripts that contain large expanded repeats and their RNA-binding proteins play an important role in the pathological course of trinucleotide-repeat-associated diseases [44•,45]. In normal cells, there are several RNA-binding proteins that specifically bind trinucleotide repeats such as CAG or CUG. In patient cells, the expanded repeats may act as a 'sink' to bind and titrate out proteins that are normally associated with other cellular mRNAs containing trinucleotide repeats, thus altering the expression of these 'downstream' mRNAs by inhibiting their processing, transport or translation [44•]. Support for this sink hypothesis comes from an unexpected finding that a short human genomic fragment containing promoter sequences and exon 1 of the Huntington's disease gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice [46••].

Conclusion and perspectives

More than 60 years ago, Hämmerling observed that the nucleus-derived information for the cap formation of *Acetabularia* traveled several centimeters and was stored at the apical tip of the stalk for several weeks. The instructions for cap formation were later shown to be mRNAs [47]. This may be the first known example of localization and translational control of mRNAs. One can envision a morphological similarity between *Acetabularia* and the neuron (cap, stalk and rhizoid versus growth cone, axon and cell body). In adult humans, each of over a trillion neurons makes connections with, on average, over a thousand target cells. It is believed that the strengthening or 'sensitization' of neuronal connections is a crude form of learning and memory [48]. How, then, can a neuron that has many axons and many synaptic connections alter the strength of only some of its synapses? It is hard to imagine that all signals received at synapses travel all the way to the nucleus — sometimes as much as several centimeters, or more than a thousand times the diameter of the cell body — and that the changes in transcriptional activity can affect just a few of a neuron's many synapses. Therefore, information that determines the strength of a synapse may be stored in a regionalized way as mRNAs at the synapse itself. Activation of the stored mRNAs through modulation of RNA-binding proteins could generate proteins important for navigation and pathfinding in response to environmental cues. Indeed, axons and growth cones have been shown to contain mRNAs [49,50••] and recent evidence suggests that synaptic transmission may require local protein synthesis within either axonal or dendritic compartments [51•].

Clearly, an understanding of the molecular basis of interactions between RNA and proteins is a prerequisite for understanding many physiological processes. Understanding how RNAs are activated through modulation of RNA-binding proteins, not only in neurons but also in other cell types, will be enhanced by knowledge of the molecular basis of RNA–protein interactions. A logical direction for future work is to resolve the relative contributions of primary sequence and structure in the recognition of RNA by proteins by determining co-crystal structures. An important step towards understanding the rules that govern RNA–protein interactions is also to continue to identify critical *cis*-acting elements and their binding proteins using molecular genetic approaches and to sort out biochemical and functional changes resulting from altered RNAs and proteins. In particular, identification of RNA-binding proteins which specifically bind *cis*-acting elements in the 3′-UTR will shed light on the molecular mechanisms of how mRNA localization, activation of translation and selected mRNA stabilization are achieved. To do so, we need to develop systems which would allow the identification of proteins that interact with relatively large regulatory elements in the 3′-UTRs, as recent methods to study RNA–protein interactions [52•–54•] appear to be suitable mostly for relatively short RNA elements.

Acknowledgements

We thank Matthias Görlach, Giovanna Musco, Stefan Grünert and Mark Bycroft for kindly providing figures for this manuscript. We are also grateful to members of the Dreyfuss laboratory, especially Sara Nakielny, Paul Eder and Victoria Pollard, for discussion and comments on this manuscript. Space limitation has precluded citation of many primary references, which are included in the reviews cited. Work in this laboratory is supported by grants from the National Institutes of Health and by the Howard Hughes Medical Institute.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. St Johnston D: **The intracellular localization of messenger RNAs.** *Cell* 1995, **81**:161–170.

2. Curtis D, Lehmann R, Zamore PD: **Translational regulation in development.** *Cell* 1995, **81**:171–178.

3. Wickens M, Kimble J, Strickland S: **Translational control of developmental decisions.** In *Translational Control.* Edited by Hershey JWB, Mathews MB, Sonenberg N. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1996:411–450.

4. Chabot B: **Directing alternative splicing: cast and scenarios.** *Trends Biochem Sci* 1996, **12**:472–477.

5. Proudfoot N: **Ending the message is not so simple.** *Cell* 1996, **87**:779–781.

6. Smith HC, Sowden MP: **Base-modification mRNA editing through deamination – the good, the bad and the unregulated.** *Trends Biochem Sci* 1996, **12**:418–424.

7. Kozak M: **An analysis of vertebrate mRNA sequences: intimations of translational control.** *J Cell Biol* 1991, **115**:887–903.

8. Foulkes NS, Sassone-Corsi P: **More is better: activators and repressors from the same gene.** *Cell* 1992, **68**:411–414.

9. Burd CG, Dreyfuss G: **Conserved structures and diversity of functions of RNA-binding proteins.** *Science* 1994, **265**:615–621.

10. Kiledjian M, Burd CG, Portman DS, Gorlach M, Dreyfuss G: **Structure and function of hnRNP proteins.** In *RNA–Protein Interactions: Frontiers in Molecular Biology.* Edited by Nagai K, Mattaj IW. Oxford: IRL Press; 1994:127–149.

11. Nagai K: **RNA–protein complexes.** *Curr Opin Struct Biol* 1996, **5**:53–61.

12. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K: **Crystal structure at 1.92Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin.** *Nature* 1994, **372**:432–438.

13. Liljas A, Garber M: **Ribosomal proteins and elongation factors.** *Curr Opin Struct Biol* 1995, **5**:721–727.

14. Bycroft M, Hubbard TJP, Proctor M, Freund SMV, Murzin AG: **The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold.** *Cell* 1997, **88**:235–242.

15. • Bycroft M, Grünert S, Murzin AG, Proctor M, St Johnston D: **NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5.** *EMBO J* 1995, **14**:3563–3571.

See annotation [16•].

16. • Kharrat A, Macias MJ, Gibson TJ, Nilges M, Pastore A: **Structure of the dsRNA binding domain of *E. coli* RNase III.** *EMBO J* 1995, **14**:3572–3584.

This paper and [15•] show, using NMR methods, that the dsRBD has an α–β–β–β–α topology in which a three-stranded anti-parallel β sheet packs on one side against the two helices. Sequence conservation and mutagenesis suggest that the amino terminus of the second helix and a nearby loop interact with dsRNA.

17. Siomi H, Matunis MJ, Michael WM, Dreyfuss G: **The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.** *Nucleic Acids Res* 1993, **21**:1193–1198.

18. • Musco G, Stier G, Joseph C, Morelli MAC, Nilges M, Gibson TJ, Pastore A: **Three-dimentional structure and stability of the KH domain: molecular insights into the fragile X syndrome.** *Cell* 1996, **85**:237–245.

The KH domain consists of a stable β–α–α–β–β–α fold. Substitution of a well-conserved hydrophobic residue located on the second helix destroys the KH fold; a mutation of this position in FMR1 leads to a severe fragile X phenotype.

19. Berglund H, Rak A, Serganov A, Garber M, Hard T: **Solution structure of the ribosomal RNA binding protein S15 from *Thermus thermophilus*.** *Nat Struct Biol* 1997, **4**:20–23.

20. • Xing Y, Guha-Thakurta D, Draper DE: **The RNA binding domain of ribosomal protein L11 is structurally similar to homeodomains.** *Nat Struct Biol* 1997, **4**:24–27.

See annotation [21•].

21. • Markus MA, Hinck AP, Huang S, Draper DE, Torchia DA: **High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA.** *Nat Struct Biol* 1997, **4**:70–77.

The RNA-binding domain of L11-C76 contains three α helices that superimpose with homeodomain α helices. The carboxy-terminal end of the third helix appears to be involved in RNA binding.

22. •• Dubnau J, Struhl G: **RNA recognition and translational regulation by a homeodomain protein.** *Nature* 1996, **379**:694–699.

A *Drosophila* homeodomain protein bicoid directly binds the 3′-UTR of the caudal mRNA and acts as a translational repressor to generate an opposing gradient of the caudal protein. This is the first demonstration of a homeodomain protein binding to RNA with specificity.

23. Portman DS, Dreyfuss G: **RNA annealing activities in HeLa nuclei.** *EMBO J* 1994, **13**:213–221.

24. Michael WM, Siomi H, Choi M, Pinol-Roma S, Nakielny S, Liu Q, Dreyfuss G: **Signal sequences that target nuclear import and nuclear export of pre-mRNA-binding proteins.** *Cold Spring Harb Symp Quant Biol* 1995, **60**:663–668.

25. Rio DC: **Splicing of pre-mRNA: mechanism, regulation and role in development.** *Curr Opin Genet Dev* 1993, **3**:574–584.

26. •• Cox JS, Walter P: **A novel mechanism for regulating activity of a transcription factor controls the unfolded protein response.** *Cell* 1996, **87**:391–404.

See annotation [27••].

27. •• Sidrauski C, Cox JS, Walter P: **tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.** *Cell* 1996, **87**:405–413.

This paper and [26••] show, using a yeast genetic screen, that a signal generated by unfolded proteins accumulated in the endoplasmic reticulum activates the pre-mRNA of the transcription factor Hac1 through a novel splicing pathway. tRNA ligase is required to finish the splicing reaction to produce $HAC1^{i}$ mRNA encoding the functional transcription factor. Interestingly, tRNA ligase as well as other proteins involved in pre-tRNA splicing have been localized near nuclear pores, suggesting that splicing of the HAC1 pre-mRNA occurs as it is exported to the cytoplasm to be translated.

28. Taylor SJ, Shalloway D: **An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis.** *Nature* 1994, **368**:867–871.

29. Fumagalli S, Totty NF, Hsuan JJ, Courtneidge SA: **A target for Src in mitosis.** *Nature* 1994, **368**:871–874.

30. • Wang LL, Richard S, Shaw AS: **p62 association with RNA is regulated by tyrosine phosphorylation.** *J Biol Chem* 1995, **270**:2010–2013.

Sam68 [70] can associate directly with *src*-family tyrosine kinases resulting in tyrosine phosphorylation of the protein. The ability of Sam68 to bind RNA is severely impaired when it is tyrosine phosphorylated, suggesting that Sam68 may play a critical role in cell cycle regulation to activate RNA.

31. Neel H, Gondran P, Weil D, Dautry F: **Regulation of pre-mRNA processing by Src.** *Curr Biol* 1995, **5**:413–422.

32. • Lee K, Haugen HS, Clegg CH, Braun RE: **Premature translation of protamine 1 mRNA causes precocious nuclear condensation and spermatid differentiation.** *Proc Natl Acad Sci USA* 1995, **92**:12451–12455.

Transgenic studies have shown that the 3′-UTR of *Prm1* mRNA is both necessary and sufficient to mediate translational repression during spermatid differentiation. This paper shows that correct temporal synthesis of *Prm1* is necessary for the transition from nucleohistones to nucleoprotamines.

33. Liu Q, Siomi H, Siomi MC, Fischer U, Zhang Y, Wan L, Dreyfuss G: **Molecular characterization of the protein products of the fragile X syndrome gene and the survival of motor neurons gene.** *Cold Spring Harb Symp Quant Biol* 1996, **61**:689–699.

34. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O *et al.*: **Diverse spermatogenic defects in human caused by Y chromosome deletions encompassing a novel RNA-binding protein gene.** *Nat Genet* 1995, **10**:383–393.

35. • Eberhart CG, Maines JZ, Wasserman SA: **Meiotic cell cycle requirement for a fly homologue of human *deleted in azoospermia*.** *Nature* 1996, **381**:783–785.

A *Drosophila* male-fertility gene, *boule*, is related to the human gene *DAZ* (for deleted in azoospermia), a candidate male fertility gene located on the Y chromosome (see [34]). Interestingly, *boule* is located on an autosome. Loss of *boule* function results in block of meiotic divisions.

36. Delbridge ML, Harry JL, Toder R, Waugh O'Neill RJ, Ma K, Chandley AC, Marshall Graves JA: **A human candidate spermatogenesis gene, *RBM1*, is conserved and amplified on the marsupial Y chromosome.** *Nat Genet* 1997, **15**:131–136.

37. Verkerk AJMH, Pieretti M, Sutcliffe J, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP *et al.*: **Identification of a gene (*FMR1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.** *Cell* 1991, **65**:905–914.

38. Oostra BA, Willems PJ: **A fragile gene.** *Bioessays* 1995, **17**:941–947.

39. De Boulle K, Verkerk AJMH, Reyniers E, Vits L, Hendrichs J, Van Roy B, Van den Bos F, De Graaff E, Oostra BA, Willems PJA: **Point mutation in the *FMR1* gene associated with fragile X mental retardation.** *Nat Genet* **3**:31–35.

40. Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G: **Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome.** *Cell* 1994, **77**:33–39.

41. Mahone M, Saffman EE, Lasko PF: **Localized *Bicaudal-C* encodes a protein containing a KH domain, the RNA binding motif of FMR1.** *EMBO J* 1995, 14:2043–2055.

42. Jones AR, Schedl T: **Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68.** *Genes Dev* 1995, 9:1491–1504.

43. Timchenko LT, Caskey CT: **Trinucleotide repeat disorders in humans: discussions of mechanisms and medical issues.** *FASEB J* 1996, 10:1589–1597.

44. • McLaughlin BA, Spencer C, Eberwine J: **CAG trinucleotide RNA repeats interact with RNA-binding proteins.** *Am J Hum Genet* 1996, 59:561–569.
Several neurological diseases are associated with an expanded CAG repeat. The authors identified cytoplasmic RNA-binding proteins that interact with CAG repeats in a tissue-specific and CAG length-dependent manner. This paper also presents a model whereby the expanded CAG repeats may act as a sink to bind and titrate away RNA-binding proteins that are normally associated with other cellular CAG-containing mRNAs.

45. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, Roberts R, Caskey CT, Swanson MS: **Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy.** *Nucleic Acids Res* 1996, 24:4407–4414.

46. •• Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, Lawton M, Trottier Y, Lehrach H, Davies SW, Bates GP: **Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice.** *Cell* 1996, 87:493–506.
The 1.9 kb human genomic fragment containing promoter sequences and exon 1 carrying expansions of approximately (CAG)130 of Huntington's disease gene has been shown to be sufficient to generate a progressive neurological phenotype that displays many of the characteristics of Huntington's disease. This work suggests that expression of expanded CAG repeats encoding polyglutamine may be sufficient to generate a mouse model of Huntington's disease which will assist in an eventual understanding of the molecular pathology of the disease.

47. Gilbert SF: *Developmental Biology*, edn 4. Sunderland, MA: Sinauer Associates, Inc.; 1994:8–10.

48. Tessier-Lavigne M, Goodman CS: **The molecular biology of axon guidance.** *Science* 1996, 274:1123–1133.

49. Van Minnen J: **RNA in the axonal domain: a new dimension in neuronal functioning?** *Histochem J* 1994, 26:377–391.

50. •• Crino PB, Eberwine J: **Molecular characterization of the dendritic growth cone: regulated mRNA transport and local protein synthesis.** *Neuron* 1996, 17:1173–1187.
This paper describes the isolation and characterization of mRNAs from single dendritic growth cones. It also shows, using single dendritic transfection, that local protein synthesis occurs in dendrites and growth cones.

51. • Kang H, Schuman EM: **A requirement for local protein synthesis in neurotrophin-finduced hippocampal synaptic plasticity.** *Science* 1996, 273:1402–1406.
The data suggest that neurotrophins stimulate the local synthesis of proteins in either axonal or dendritic compartments that are required for the induction of synaptic enhancement. The newly synthesized proteins may act locally to enhance postsynaptic responsiveness or may communicate with the presynaptic terminal to increase neurotransmitter release.

52. • Sen-Gupta DB, Zhang B, Kraemer B, Prochart P, Fields S, Wickens M: **A three-hybrid system to detect RNA-protein interactions *in vivo*.** *Proc Natl Acad Sci USA* 1996, 93:8496–8501.
A three-hybrid system is a variation of the yeast two-hybrid technology; it allows selection for specific RNA-binding proteins using the following features. A vector expressing the lexA DNA-binding domain fused to the phage MS-2 protein is integrated into the yeast genome. A vector expressing a hybrid RNA that contains the MS-2 protein-binding site and the target RNA sequence of interest and a cDNA library fused to the Gal4 activation domain are introduced into the yeast. Expression of a fusion protein that interacts with the target RNA sequence will result in activation of transcription from promoters that bind the lexA–MS2 fusion protein.

53. • Harada K, Martin SS, Frankel AD: **Selection of RNA-binding peptides *in vivo*.** *Nature* 1996, 380:175–179.
In the two-plasmid system used in this paper, the antiterminator protein N is expressed under the control of a *tac* promoter and β-galactosidase is also expressed under control of a *tac* promoter from the second plasmid with the N-utilization (nut) and termination sites upstream of *lacZ*. Because the interaction of N with nut is mediated by an arginine-rich region of N, it is anticipated that the interaction may be replaced by other arginine-rich peptide–RNA interactions.

54. • Jain C, Belasco JG: **A structural model for the HIV-1 Rev–RRE complex deduced from altered-specificity Rev variants isolated by a rapid genetic strategy.** *Cell* 1996, 87:115–125.
The method used in this paper is to subject *lacZ* expression to translational control by a heterologous RNA-binding protein by inserting a binding site for that protein just upstream of an RNA element important for translation initiation (i.e. the Shine–Dalgarno element). Binding by the RNA-binding protein in *E. coli* would sterically hinder ribosome binding and reduce β-galactosidase production.

55. • McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL: **The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.** *Nature* 1997, 385:357–361.
This paper shows that the carboxy-terminal domain of the RNA polymerase II large subunit is required for efficient RNA processing such as splicing and polyadenylation. The carboxy-terminal domain appears to associate directly with splicing factors and 3′ processing factors. The data suggest that splicing and polyadenylation are not only coupled to one another but also to transcription.

56. •• Takagaki Y, Seipelt RL, Peterson ML, Manley JL: **The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation.** *Cell* 1996, 87:941–952.
Experiments described in this paper show that overexpression of one of the polyadenylation factors (CstF-64) in a B cell line increases the concentration of intact CstF and is sufficient to switch IgM heavy chain expression from the membrane-bound form to the secreted form.

57. Dietz HC, Kendzior RJ Jr: **Maintenance of an open reading frame as an additional level of scrutiny during splice site selection.** *Nat Genet* 1994, 8:183–188.

58. Gorlach M, Wittekind M, Beckman RA, Mueller L, Dreyfuss G: **Interaction of the RNA-binding domain of the hnRNP C proteins with RNA.** *EMBO J* 1992, 11:3289–3295.

59. Bevilacqua PC, Cech TR: **Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR.** *Biochemistry* 1996, 35:9983–9994.

60. Draper BW, Mello CC, Bowerman B, Hardin J, Priess JR: **Mex-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos.** *Cell* 1996, 87:205–216.

61. Urlaub H, Kruft V, Bischof O, Muller E-C, Wittmann-Liebold B: **Protein–rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies.** *EMBO J* 1995, 14:4578–4588.

62. Schwabe JWR, Klug A: **Zinc mining for protein domains.** *Nat Struct Biol* 1994, 1:345–349.

63. Ghisolfi L, Joseph G, Amalric F, Erard M: **The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties.** *J Biol Chem* 1992, 267:2955–2959.

64. Puglisi JD, Chen L, Blanchard S, Frankel AD: **Solution structure of a bovine immunodeficiency virus Tat–TAR peptide–RNA complex.** *Science* 1995, 270:1200–1203.

65. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR: **Alpha helix–RNA major groove recognition in an HIV-1 Rev peptide–RRE RNA complex.** *Science* 1996, 273:1547–1551.

66. Lindahl M, Svensson LA, Liljas A, Sedelnikova SE, Eliseikina IA, Fomenkova NP, Nevskaya N, Nikonov SV, Garber MB, Muranova TA *et al.*: **Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*.** *EMBO J* 1994, 13:1249–1254.

67. Predki PF, Nayak M, Gottlieb MBC, Regan L: **Dissecting RNA–protein interactions: RNA–RNA recognition by Rop.** *Cell* 1995, 80:41–50.

68. Navaratnam N, Bhattacharya S, Fujino T, Patel D, Jarmuz AL, Scott J: **Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site.** *Cell* 1995, 81:187–195.

69. Hentze MW: **Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains?** *Trends Biochem Sci* 1994, 19:101–103.

70. Lock P, Fumagali S, Polakis P, McCormick F, Courtneidge SA: **The human p62 cDNA encodes Sam68 and not the RasGAP-associated p62 protein.** *Cell* 1996, 84:23–24.