Zipcodes and postage stamps: mRNA localisation signals and their trans-acting binding proteins

Hervé Chabanon, Ian Mickleburgh and John Hesketh

Received (in revised form): 23rd August 2004

Abstract

Messenger RNA (mRNA) localisation is a widespread mechanism within eukaryotic cells that provides local synthesis of proteins close to where they function. In general, this mRNA targeting involves the cytoskeleton and signals within the 3' untranslated region (3'UTR) of the transcript. In this paper, the authors review what is known of the nature of the localisation signals and the proteins that interact with them in animal cells. Specific examples are selected to illustrate the emerging pattern of how these signals are formed by the mRNA and the key RNA-binding proteins. The signals are usually restricted to relatively short regions of the 3'UTR, but their precise nature varies, with both sequence and structure playing key roles. Repeat motifs and functional redundancy also appear as common features of these signals. The trans-acting factors involved in localisation include proteins having other roles in nuclear events, proteins that shuttle between the nucleus and the cytoplasm and translational factors. In addition, there is evidence of homology among these proteins and the mechanisms of localisation across eukaryotic species.

Keywords: 3' untranslated region, RNA-binding protein, targeting, messenger RNA, localisation

MESSENGER RNA LOCALISATION

Eukaryotic organisms have developed a variety of mechanisms to achieve subcellular targeting of newly synthesised proteins to their site of function. In addition to the well-documented signal polypeptide sequences within proteins, such mechanisms include localisation of messenger RNAs (mRNAs) to particular cytoplasmic sites. mRNA trafficking or localisation produces a spatial organisation of the synthesis of specific proteins within cells close to where they function.1-4 This occurs in a range of cell types and organisms from yeast to Drosophila to mammals and plants. This paper focuses on the mechanisms of cytoplasmic mRNA localisation in eukaryotic animal cells. Separate but related topics — such as, for instance, association of mRNAs with the rough endoplasmic reticulum,5 nucleolar and subnucleolar localisation of ribosomal RNAs and heterogeneous nuclear RNAs6,7 or RNA targeting to organelles (eg mitochondria and chloroplasts)8-10 — will not be developed here. The aim is to review emerging patterns in the cis-acting signals, or 'zipcodes', that determine cytoplasmic mRNA localisation and the trans-acting factors that bind to them so as to highlight the opportunities for genomic and proteomic approaches to contribute to this field.

In Saccharomyces cerevisiae, there is localisation of specific transcripts, such as Ash1, in the budding cytoplasm.4 In early Drosophila embryos, several mRNAs are localised to the anterior or posterior pole (eg bicoid mRNA to the anterior pole and nanos mRNA to the posterior pole), and this is absolutely vital for determining the gradient of the encoded protein morphogens and subsequent developmental fate.11 In highly polarised cells, such as neurons, increasing numbers of mRNAs have been found to
be localised either to dendrites or to the synaptic endings of axons. This requires transport of mRNAs and components of the protein synthetic machinery over considerable cellular distances and produces protein synthesis both close to synaptic endings and in dendrites. This contributes to the spatial polarisation of these highly complex cells. Similarly, myelin basic protein (MBP) mRNA is transported down the long cell processes of oligodendroglia and translation is repressed until the messenger is localised. In both neurones and oligodendroglia, there is evidence that the transport of the mRNAs takes place in granules.

Even in less overtly polarised cells, such as fibroblasts and myoblasts, certain mRNAs are not distributed throughout the cytoplasm but are localised to specific subcellular sites. For example, β-actin mRNA is found in the cell periphery under conditions when actin synthesis in this area is increased. By contrast, certain mRNAs are found associated with the cytoskeleton and localised in the perinuclear cytoplasm. Whereas β-actin mRNA is found in the cell periphery, γ-actin mRNA is found around the nucleus, similarly, creatine kinase isoform mRNAs B and M are found in the perinuclear and peripheral cytoplasm, respectively. Perinuclear mRNAs also include those encoding the nuclear transcription factors MYC and FOS, as well as the mRNA encoding metallothionein isoform 1 (MT1), which is normally found in the cytoplasm but which is imported into the nucleus during the G1/S transition of the cell cycle. Disruption of mRNA localisation causes impaired protein localisation such that β-actin distribution and cell motility is affected or MT1 relocalisation to the nucleus prevented and cell function impaired.

Therefore, there is considerable evidence that mRNA targeting has important functions in determining the final location of the encoded protein. The physiological basis of mRNA localisation varies. It can enhance the efficacy of protein targeting (eg β-actin, MT1), allow the sorting of closely related protein isoforms (eg actin, creatine kinase) or promote the association of the protein with a specific cell structure (eg cyclin B with the mitotic spindle). In addition, mRNA localisation can restrict protein activity or concentration to a particular region of the cell, such as, for example, in oligodendroglia where localisation of MBP mRNA limits the synthesis of the MBP protein to regions where myelin should be laid down or in Drosophila embryos where localisation of bicoid mRNA generates a gradient of the morphogen bicoid.

**ROLE OF THE 3’ UNTRANSLATED REGION**

mRNA localisation can be produced by differential stability of the transcripts in different subcellular locations or by active transport mechanisms that target them to different destinations. To date, the evidence suggests that, in the vast majority of cases, localisation of mRNAs is achieved by targeting mechanisms involving transport and anchoring as is schematically illustrated in Figure 1. Such targeting is mostly brought about by specific mechanisms that involve components of the cytoskeleton and signals within the 3’ untranslated region (3’UTR) of the mRNA. Garken and yemanuclein-α mRNAs in Drosophila, and the yeast Ash1 mRNA, represent exceptions to this rule; as their localisation elements also involve the 5’UTR or coding region. In Drosophila, mutations in regions corresponding to 3’UTRs lead to loss of mRNA localisation and aberrant development. Moreover, studies with transfected mammalian cells have shown that addition of a 3’UTR from localised mRNAs (eg β-actin, c-myc) to a reporter transcript leads to localisation of that reporter. Similarly, removal of the 3’UTR causes loss of localisation of the mRNA. Thus, for the majority of localised mRNAs identified so far, both
in somatic and germ cells, localisation depends on a cis-acting element(s) that resides in the 3′UTR. This mRNA localisation, via a signal within the 3′UTR of the mRNA itself, is distinct from the association of ribosome–mRNA complexes with the endoplasmic reticulum through the signal sequence present within the nascent polypeptide chain.

More detailed studies using deletion analysis and mutagenesis show that the whole 3′UTR is not necessary for localisation, but rather that one or more restricted regions or localisation elements is sufficient. Furthermore, it is the combination of these cis-acting signals and the trans-acting factors that determine if a particular mRNA will be localised in the cytoplasm and where it will be translated. The approach in this paper has been selective but, where possible, consistent features concerning the nature of the localisation signals (LSs) and the classes of proteins that interact with them are highlighted.

**THE CIS-ACTING mRNA LOCALISATION SIGNALS**

From present data, it is difficult to define exactly what features make up LSs (or zipcodes) because these are highly variable in length, structure and complexity (Table 1 and Figure 2). Several general principles are now emerging. First, LSs can be either short single motifs or multicomponent
Table 1: RNA–protein interactions involved in mRNA localisation

<table>
<thead>
<tr>
<th>mRNA (cell/organism)</th>
<th>Localisation</th>
<th>Cis-acting signal</th>
<th>Trans-acting factors</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-actin (Chicken embryo fibroblasts)</td>
<td>Leading edge of the lamella</td>
<td>54 nt zipcode (43 nt zipcode-like)</td>
<td>Tandem repeat ACACCC</td>
<td>ZBP1, ZBP2 eEF1α for anchoring 22, 29, 30, 31</td>
</tr>
<tr>
<td>Myelin basic protein, MBP (mouse oligodendrocytes)</td>
<td>Transport along processes. Localisation in myelin compartment</td>
<td>Modular RTS or A2RE RLR (Region 1131–1473)</td>
<td>11 nt sequence = A2RE11. Secondary structure</td>
<td>hRNP A2 13, 32</td>
</tr>
<tr>
<td>Vimentin (3T3 fibroblasts, CHO cells)</td>
<td>Perinuclear cytoplasm</td>
<td>100 nt</td>
<td>Secondary structure?</td>
<td>HAX-1, eEF1γ, hRIP 33, 34</td>
</tr>
<tr>
<td>Metallothionein-I (CHO cells)</td>
<td>Perinuclear cytoplasm</td>
<td>40 nt</td>
<td>Sequence repeat CACC + structure?</td>
<td>eEF1α! 35, 36</td>
</tr>
<tr>
<td>c-myc (fibroblasts)</td>
<td>Perinuclear cytoplasm</td>
<td>86 nt</td>
<td>Structure? + conserved AUUUA</td>
<td>N/a 19</td>
</tr>
<tr>
<td>Bicoid (Drosophila oocyte/embryo)</td>
<td>Anterior</td>
<td>625 nt — modular 53 nt BLE1 helical domains (stems III, IV, V)</td>
<td>Secondary and higher order structure</td>
<td>Staufen 37–39, 40, 41, 42, 44, 44</td>
</tr>
<tr>
<td>Vg1 (Xenopus oocyte)</td>
<td>Vegetal cortex</td>
<td>340 nt</td>
<td>Clusters of VM1 (or E1) (consensus 5’-YYUCU-3’) and E2 motifs (CAC-containing motifs)</td>
<td>Vera/Vg1RBP (ZBP1 like) 45–52</td>
</tr>
<tr>
<td>VegT (Xenopus oocyte)</td>
<td>Posterior</td>
<td>525 nt</td>
<td></td>
<td>VgRBP60/hnRNP I</td>
</tr>
<tr>
<td>Nanos (Drosophila oocyte/embryo)</td>
<td>Posterior (germ plasm)</td>
<td>547 nt</td>
<td>N/a</td>
<td>p75 53–55</td>
</tr>
<tr>
<td>Oskar (Drosophila oocyte/embryo)</td>
<td>Posterior</td>
<td>Modular. Multiple elements of 100–200 nt</td>
<td>N/a</td>
<td>56</td>
</tr>
<tr>
<td>K10 (Drosophila oocyte/embryo)</td>
<td>Anterior</td>
<td>TLS = 44 nt region</td>
<td>Single stem–loop</td>
<td>57</td>
</tr>
<tr>
<td>Gurken (Drosophila oocyte/embryo)</td>
<td>Dorsal–anterior</td>
<td>Modular. Different elements in 5’ UTR, CDS and 3’ UTR</td>
<td>N/a</td>
<td>Squid 24, 25, 58</td>
</tr>
<tr>
<td>Hairy (Drosophila oocyte/embryo)</td>
<td>Blastoderm apical cytoplasm</td>
<td>HLE (region 1281–1406)</td>
<td>SL1/SL2a. Secondary and higher order structure</td>
<td>59</td>
</tr>
<tr>
<td>Ash1 (budding yeast)</td>
<td>Daughter cell</td>
<td>Several LE E1, E2a, E2b (CDS) E3 (end CDS – 3’ UTR)</td>
<td>Secondary structure</td>
<td>27, 28</td>
</tr>
</tbody>
</table>

Abbreviations used: A2RE, hnRNP A2 response element; BLE1, bicoid LE1; CDS, coding sequence; CHO, Chinese hamster ovary; eEF, eukaryotic elongation factor; HAX-1, HS1-associated protein X-I; HLE, hairy LE; hnRNP, heterogeneous nuclear ribonucleoprotein; hRIP, human rev-interacting protein; LE, localisation element; ME, minimal element; nt(s), nucleotide(s); RBP, RNA-binding protein; RLR, RNA localisation region; RTS, RNA transport sequence; SL1/2a, stem–loop 1/2a; TLS, transport/localisation sequence; UTR, untranslated region; VM1, Vg1 motif 1; ZBP, zipcode-binding protein. In the consensus sequence 5’-YYUCU-3’, Y is U or C.
units spanning over hundreds of nucleotides. Second, different sub-elements, overlapping, in close proximity or spread throughout the transcript, can either direct discrete steps of the localisation pathway or can be redundant for the localisation function.

**Multiplicity of the LS**

*Oskar* and *bicoid* mRNAs, which are both localised in the *Drosophila* embryo, provide two well-defined examples of modular organisation of LSs. Deletion analysis of the *oskar* 3’UTR identified several 100–200 nucleotide (nt) regions that act independently in distinct steps of the localisation programme that directs the transcript to the posterior pole of the oocyte.56 In *bicoid* mRNA, a 53 nt region, *bicoid* localisation element 1 (BLE1), only supports early stages of localisation while later stages require additional elements.37 In particular, anchoring of the transcript at the anterior pole involves interactions of the double-stranded RNA-binding protein staufen with different helical domains spread over the *bicoid* LS (Figure 2).38,39,44 A combinatorial organisation of LEs is also found in the MBP mRNA. First assembled into granules in the perikaryon, this messenger is then transported along the processes by way of a 21 nt sequence called RNA transport sequence (RTS), while a further element, the RNA localisation region (RLR), mediates its specific localisation.13 Interestingly, other transcripts contain a sequence homologous to the RTS — eg the MAP2A mRNA in hippocampal neurones or the protamine-2 mRNA in spermatocytes — and these were shown to be transported in oligodendroglia, suggesting that a common recognition event occurs in different cell systems for different mRNAs.13,32 It is thus conceivable that there are some targeting elements, such as the RTS, that are common to a range of mRNAs and that the combination of different sub-elements would allow RNAs with ultimately distinct destinations to use similar transport mechanisms for common intermediate steps in their respective pathways. Another example of a combination of signals is given by *Xsat-2* mRNA in *Xenopus* oocytes. Targeting of this messenger to the mitochondrial cloud by the early (or METRO) pathway depends on a general element termed the mitochondrial cloud localisation element (MCLE); however, subsequent association of *Xsat-2* mRNA with the germinal granules inside the cloud requires the germinal granule localisation element (GGLE) as an additional specific element present in the 3’UTR.60,61

The existence of multiple, partially redundant, elements has been observed during the mapping of the *cis*-acting signal in a number of localised mRNAs. This suggests that some degree of functional redundancy may be another characteristic of mRNA LSs. In the *β-actin* mRNA, a 43 nt region with a sequence reminiscent of the fully active 54 nt zipcode presents only a partial localisation activity;22 however, it is more common to find several different elements with each having only a partial activity on their own. The need for multiple elements to be associated to achieve optimal localisation efficiency suggests that some cooperative interactions between them are required to recruit the localisation machinery. For example, deletion analysis of the 3’UTR and/or insertion of different combinations of 3’UTR sequences into a heterologous mRNA revealed just such a synergistic action between sub-elements present in *nanos* or *fatvg* mRNA.53,54,62,63 In *bicoid* and *Vg1* mRNAs, small deletions within BLE1 and the whole *Vg1* LS, respectively, were unable to disrupt localisation.37,43 again suggesting some functional redundancy. Moreover, the BLE1-mediated pathway does not seem to be essential, since another redundant pathway with different sequence requirements was identified that allows *bicoid* mRNA localisation.40

Finally, for *bicoid*, *nanos*, *hairy* and *Vg1* mRNAs, strong localisation efficiency can be achieved by multimerisation of one sub-element that shows only a weak
Figure 2: Types of cis-acting mRNA localisation signals and their binding proteins. Both structure- and sequence-based recognition motifs have been described to characterise mRNA LSs. (A) The bicoid LS is made up of a complex structure with a modular organisation. The sequential steps of the localisation programme involve distinct, partially overlapping, regions of the signal (BLE1 for early localisation stages and staufen-binding domains for the mRNA anchoring). Contrasting with the complex structure of bicoid LS, a single 44 nt stem-loop, termed the transport/localisation sequence (TLS), forms the localisation element of K10 mRNA. (B) MBP, β-actin, Vg1 and VegT are well-described examples of mRNAs presenting a LS comprising a particular sequence. The RTS of MBP mRNA corresponds to the hnRNP A2 response element (A2RE). A tandem repeat of an ACACCC motif, which binds ZBP1, is essential for the localisation activity of the β-actin mRNA zipcode. The translation factor, eEF1α, may also intervene in the localisation pathway to anchor the transcript on the actin cytoskeleton. Finally, overlapping clusters of VM1 and E2 sites define the LSs of Vg1 and VegT mRNAs. See Table 1 for abbreviations and references.
(nanos, hairy) or no (Vg1, BLE1) localisation activity on its own, suggesting the need for reiterated sites for the binding of a localisation factor(s).\textsuperscript{37,45,55,59,62} Overall, it appears that, in many cases, LSs are made up of multiple elements, acting either independently or in a concerted manner. This emphasises the potential complexity of the RNA–protein interactions involved in mRNA localisation and the need to define the precise nature of LSs as key factors in the characterisation of these interactions and the understanding of localisation mechanisms.

**Structure or sequence**

Numerous cis-acting elements that regulate mRNA translation and stability have been defined and can be either sequences (eg the AU-rich elements that give mRNA instability\textsuperscript{64} or the element that regulates lipoxygenase mRNA translation\textsuperscript{65}) or structural features (eg the iron-regulatory element [IRE] stem–loop involved in the binding of the iron-responsive protein\textsuperscript{66} or the SECIS [SElenoCysteine Insertion Sequence] hairpin structure responsible for the recoding of the stop codon UGA as the codon for selenocysteine insertion during the translation of selenoprotein mRNAs\textsuperscript{67}). The situation with regard to LSs is far less clear but there is evidence for the importance of both structural- and sequence-based recognition motifs (Table 1 and Figure 2).

Localisation of *bicoid*, K10 and hairy mRNAs in *Drosophila*, and *Ash1* mRNA in the budding yeast, has been shown to depend on the formation of specific structural motifs (Table 1). More precisely, a single AU-rich stem–loop of 44 nt is responsible for the transport and the localisation of the K10 transcripts to the oocyte’s anterior cortex\textsuperscript{57} (Figure 2). The localisation signal in *bicoid* mRNA, initially defined as a 625 nt region by genetic complementation, is predicted to fold into a complex secondary structure with five large stems (I–V) and is well conserved between various *Drosophila* species.\textsuperscript{41,68,69} Stem–loops IV and V were shown to support the recognition event responsible for the early transport of the mRNA into the embryo.\textsuperscript{40} Extensive site-directed mutagenesis of the BLE1–containing stem–loop V revealed the requirements for two particular domains: a recognition domain formed by two base pairs present between two dinucleotide bulges and the terminal stem domain.\textsuperscript{42} A similar approach highlighted the double-stranded regions as the major determinant for the activity of the K10 transport/localisation sequence (TLS) and *hairy* stem–loop 1/2a Es.\textsuperscript{57,59} Yet, the secondary structure by itself does not seem to be sufficient to provide binding specificity. Compensatory base changes that retain base pairing in the recognition domain of *bicoid* and *hairy* stem–loops 1/2a suggest that higher order structures are involved as optimal localisation efficiency can be achieved only when identity of the base pairs is conserved.\textsuperscript{42,59} In addition, for *bicoid* mRNA, a loop–loop interaction between *bicoid* transcripts via single-stranded nucleotides present in stem–loop III seems to participate in the formation of staufen-containing ribonucleoprotein (RNP) particles required for the messenger localisation.\textsuperscript{33,44} To date, this is a unique example of the critical role of a quaternary structure in the localisation complex.

By contrast, in other mRNAs, the localisation signals appear to be characterised by specific sequences (Table 1). An 11 nt sequence in the MBP mRNA (referred to as A2RE11) or heterogeneous nuclear ribonucleoprotein [hnRNP] A2 response element) has been identified as the minimum requirement both for the binding of the transport of the messenger (Figure 2).\textsuperscript{32} Interestingly, the capacity of A2RE-like sequences present in other localised mRNAs to bind to this protein could be correlated with the transport efficiency of these transcripts in oligodendrocytes.\textsuperscript{32} Several studies suggest that LSs can comprise repeat sequences. For example,
a tandem repeat of an ACACCC sequence in the zipcode of the chicken β-actin mRNA is essential for the binding of zipcode-binding protein 1 (ZBP1) and the transcript localisation to the lamellipodia. In Xenopus, localisation of VgI and VegT mRNAs to the vegetal cortex of the oocyte depends on repetitions of CAC-containing motifs (or E2 motifs) present in the LEs that mediate the binding of Vera/Vg1 ribonucleotide binding protein (RBP), a frog homologue of ZBP1. Using a computer-assisted search (REPFIND program), Betley et al. recently found clusters of CAC-containing motifs in the LEs of nine out of ten RNAs localised to the vegetal cortex of Xenopus oocytes. They also identified new vegetally localised mRNAs on the basis of their content in such clusters. Although necessary, the cluster of E2 motifs in VgI and VegT LEs is not sufficient to promote localisation and needs to be associated with a second overlapping cluster of VgI motif 1 (VM1) sites. The VM1 sites, with a consensus sequence recently defined as 5'YYYUCU-3', bind VgRBP60, a homologue of the mammalian hnRNP I. Interestingly, the proximity of E2 and VM1 sites within the LE is believed to generate an interaction between the two binding proteins that could be critical for the localisation process. Strikingly, in situ hybridisation experiments carried out in the authors’ laboratory showed that an 11 nt sequence containing a repetition of a CACC motif present in the 3'UTR is necessary for perinuclear localisation of MT1 mRNA; this repeat confers some localisation properties when introduced into the 3'UTR of a heterologous mRNA. The clustering of small reiterated motifs in a number of LEs supports the view that they represent a novel type of key recognition element involved in recruitment of the trans-acting factors for localisation. Such repeated motifs may be part of a common strategy, conserved during evolution, for the targeting of distinct mRNAs.

Present knowledge of localised mRNAs and their signals comes from studies of particular transcripts. As yet, the use of genomic approaches to identify other mRNAs containing such signals has not been possible. Further insights into the definition of LEs will be necessary before bioinformatic analysis of genome sequence data can be used to predict their existence in 3'UTR sequences. Currently, there are several immediate challenges. In the case of sequence-based recognition elements, the presence in one transcript of a motif critical for the localisation of other mRNAs does not automatically mean that this motif is functional in this particular transcript. For example, although the 3'UTRs of Fatvg and Xvelo1 mRNAs contain both E2 and VM1 sites, it turns out that localisation of these two mRNAs to the vegetal cortex of Xenopus oocytes occurs via a different LE. For structure-based recognition elements, predictions of RNA secondary structure based on computer algorithms, like the MFOLD program, and/or by phylogenetic sequence comparison have proved useful. In the case of bicoid 3'UTR, chemical and enzymatic probing have confirmed the main structural features of the LS; however, the bioinformatic tools dedicated to the prediction of RNA folding do not allow the determination of specific local structures and non-canonical base pairings found in RNAs, the number of which reported in known RNA structures is growing. The requirement for higher order structures in mRNA signal recognition by the localisation machinery, as suggested by the observations mentioned above, makes it very difficult to determine the existence of cis-acting LEs by analysing the primary sequence or predicting the secondary structure of a given mRNA. Only one stem–loop predicted to form within the LS of the Drosophila orb mRNA was found as an apparent counterpart of the K10 TLS. Finally, protein binding to the mRNA LEs may
modify the transcript structure, influence the formation of different recognition domains and affect localisation. In general, the use of computer prediction is still limited and experimental structural determination remains important.

THE TRANS-ACTING FACTORS THAT BIND TO LOCALISED mRNAs

Localisation depends not only on the mRNA signal but also on the proteins that interact with these regions to form the critical localisation complex. There is now evidence that mRNAs are found in protein-containing cytoplasmic granules involved in their localisation and translation. These granules are likely to be distinct from the processing bodies (P bodies) that are associated with mRNA decapping and decay in yeast. It is possible that the RNA transport granules contain not only destination-specific proteins (ZBPs) but also more general components that act as ‘adaptor’ proteins in the RNP complex. An example of the latter is the double-stranded RNA (dsRNA)-binding protein staufen which has the ability to bind to a variety of mRNAs. Originally discovered in Drosophila, staufen has been studied in detail. It binds directly to oskar, bicoid and prospero mRNAs. Of its five dsRNA-binding domains (dsRBDs; approximately 70 amino acids in length and containing several conserved basic and hydrophobic residues), only domains 1, 3 and 4 show binding properties in vitro. Domain 2 is required for microtubule-dependent localisation of oskar mRNA and domain 5 is required for the interaction with Miranda that allows actin-dependent localisation of prospero mRNA. A mammalian homologue of staufen has been found and the role of this protein in mRNA localisation in both Drosophila and mammalian neurones has been reviewed thoroughly elsewhere. Knowledge of the other proteins involved in localisation is still limited but candidate proteins have been identified (Table 1). Although these candidates are unlikely to represent the full complement of trans-acting factors, certain patterns in the types of proteins involved have emerged. The characteristics of these proteins are that they can be either dsRNA- or single-stranded RNA-binding proteins, proteins that shuttle between the nucleus and the cytoplasm or proteins previously identified as translation factors.

Nuclear proteins that shuttle into the cytoplasm

The hnRNPs are a large family of predominantly nuclear proteins associated with mRNAs. A subset of these has been shown to shuttle between the nucleus and the cytoplasm. hnRNP I shares 87 per cent amino acid sequence identity with the Xenopus VgRBP60 protein, which binds the VMI1 element of the Vg1 mRNA 3’UTR and co-localises with this messenger. The Drosophila Squid protein is a homologue of hnRNP A1, and this localises gurken mRNA to the anterior dorsal cortex of the oocyte through binding to two distinct regions of the 3’UTR. hnRNP A2 has been shown to mediate the localisation of MBP mRNA by binding to A2RE. Although mostly located in the nucleus, hnRNP A2 is also found in cytoplasmic granules. hnRNP A2 has higher affinity for A2REs than other hnRNPs and this involves its RNA recognition motif (RRM) domains that consist of an amino acid sequence of 90–100 residues with two short consensus sequences separated by mainly hydrophobic amino acids binding in concert. Spliceosomes extracted from human cells mediate the formation of a multiprotein complex 20–24 nt upstream of an exon–exon junction complex (EJC).

Staufen is a well-defined example of a general protein factor acting in localisation of several mRNAs
Removal of the first exon–exon junction of oskar results in delocalisation of the mRNA from the posterior pole, suggesting that splicing is a key requirement for oskar mRNA localisation as well as the 3’UTR LE.98 One of the prime examples of a mRNA transported to the cell periphery is β-actin. Its 54 nt zipcode has been shown to bind directly to the proteins ZBP129 and ZBP2.30 The 92 kilodalton (kD) ZBP2, isolated by Gu et al.,30 is a homologue of the nuclear RNA splicing factor KSRP, which binds an intronic splicing enhancer element.99 Both of these proteins possess four K homology (KH) domains which bind single-stranded nucleic acids in vitro and which were first identified in the human hnRNP K through which they bind RNAs.100,101 ZBP2 also has a 47 amino acid region not present in KSRP. Although ZBP2 is predominantly nuclear, it has been shown to co-localise with the β-actin mRNA at the leading edges of chicken embryonic fibroblasts, suggesting an association with this messenger. The MAP2 mRNA trans-acting protein 1 (MARTA1) also shares a high degree of sequence identity (98 per cent) with the human KSRP.102 ZBP1 is a 68 kD protein mainly found in the cytoplasm; it also possesses nuclear localisation and export signals, suggesting that it may be present in the nucleus transiently.29 This protein belongs to an evolutionarily conserved family including the Xenopus protein Vera/Vg1RBP which binds Vg1 and VgT mRNAs.47,50 Members of this family possess two RRM domains at their amino terminal and four KH domains near their carboxy terminal. Farina et al.103 have demonstrated that the KH domains 3 and 4 of ZBP1 are essential for binding the β-actin mRNA zipcode, for the association with the actin cytoskeleton and for the formation of cytoplasmic granules.

The picture emerging from these studies is that proteins involved in localisation are able to shuttle between the nucleus and the cytoplasm. This leads to the concept of mRNAs becoming ‘tagged’ for transport and localisation by the binding of specific proteins in the nucleus and during mRNA processing and export.

Translation factors implicated in mRNA localisation

Palacios et al.104 have shown that the translation initiation factor eIF4AIII interacts directly with the dimer Magoh–Y14 and is recruited to the EJC in the nucleus (both in human-derived cells and in Drosophila oocytes). Using a nonsense-mediated decay (NMD) assay combined with RNA interference (RNAi), both eIF4AIH and the mRNA localisation protein Barentsz were shown to be required for the NMD process in HeLa cells.104 When exported into the cytoplasm, the EJC binds to Barentsz through eIF4AIH.105 As Barentsz is essential for oskar mRNA localisation, eIF4AIH must also be necessary for this localisation since it forms a link between Barentsz and the Magoh–Y14 dimer in the EJC.104,105

The translation elongation factor eEF1α is the second most abundant protein in eukaryotes and is, therefore, in vast molar excess to other proteins involved in translation.106 This protein has been shown to be associated with actin filaments and microtubules in cells from various organisms.106 eEF1α has been shown to co-localise with poly(A) RNAs and ribosomes on the actin cytoskeleton in fibroblasts.107 Liu et al.31 have demonstrated that β-actin mRNA is bound by eEF1α when in complex with F-actin in vitro and that β-actin mRNA and eEF1α co-localise in cell protrusions in vivo. Although this interaction was found to be of high affinity, the 3’UTR was not required and the anti-sense sequence was able to compete for binding. These data suggest that eEF1α may bind to a secondary structure element. It has been proposed that binding of cEF1α to a possible stem–loop structure anchors mRNAs to the actin cytoskeleton.31 Using competitive RNA-binding assays, the authors have demonstrated that eEF1α binds...
specifically to the perinuclear localisation element in the 3' UTR of MT1 mRNA. These data implicate eEF1α in anchoring certain mRNAs in specific areas of cells through the actin cytoskeleton.

The poly(A)-binding protein (PABP), which possesses four amino terminal RRM domains, has been reported to be involved in the localisation of vasopressin (VP) mRNA in rat nerve cell processes. Although this protein is well characterised as binding the poly(A) tail of mRNAs, this study suggests that it also specifically binds to the dendritic localiser sequence (DLS) of VP mRNA. In addition, Arn et al. have isolated PABP as a component of a multiprotein complex binding to the bicoid localisation sequence in Drosophila.

The vimentin mRNA contains a highly conserved 100 nt element within its 3' UTR, which forms a bifurcated stem-loop that is necessary for its localisation to the perinuclear cytoplasm. A 70 nt sequence from this region was used by Al-Maghrebi et al. in a yeast triple-hybrid screen to identify binding proteins. The proteins identified were hRIP (32 kD), HAX-1 (35 kD) and the translation elongation factor eEF1γ (50 kD). HAX-1 showed sequence-specific binding, whereas eEF1γ and hRIP were shown to be less specific. The role of eEF1γ in translation is unclear because, although it can stimulate the recycling of eEF1α-GDP to eEF1α-GTP by forming a complex with eEF1β and δ, it is not essential for this process. The hydrophobic tail of eEF1γ seems to be involved in interactions with membranes and the cytoskeleton, so it could anchor mRNAs to these structures. HAX-1 also possesses a hydrophobic domain and has been shown to be associated with membranes throughout the cell. Moreover, vimentin mRNA appears to co-localise with the cytoskeleton. Al-Maghrebi et al. suggest that perinuclear-localised mRNAs may be bound by proteins anchored to the cytoskeleton or nuclear membrane as they leave the nucleus, to prevent diffusion throughout the cytoplasm.

After leaving the nucleus, certain mRNAs remain perinuclear whereas others are destined to be localised in other parts of the cell. Localisation to the cell periphery requires active transport along the cytoskeleton and then anchoring at the destination, but perinuclear mRNAs may only need to be anchored near the nucleus after export from it (illustrated in Figure 1). Thus, one might expect transported and perinuclear mRNAs to bind different complements of proteins in which some of those involved in the anchoring process may be common; an example of the latter may be eEF1α.

CONCLUSIONS AND PERSPECTIVES

mRNA LSs, or zipcodes, are formed by a discrete region or regions, most often in the 3' UTR. These signals bind a series of specific proteins to form RNP localisation complexes that, in some experimental systems, have been detected in the cytoplasm as RNA-containing granules. Presumably, the formation and exact composition of these complexes determine the transport, sorting and localisation of the transcripts in the cytoplasm. These events are complex and localisation involves several steps — sorting and transport stages, and an anchoring of the mRNA at the final destination. It is beginning to emerge that this requires multiple signals and proteins, as well as links with other post-transcriptional events such as splicing, export and translation; this is shown schematically in Figure 1. The identification of the trans-acting proteins is showing that both ‘destination-specific’ transport and anchoring proteins as well as general ‘adaptor’ proteins (eg staufen) are likely to make up these complexes. It is also likely that protein binding is sequential, starting within the nucleus, and that such nuclear events influence subsequent events in the cytoplasm. For example, splicing influences localisation in the case of oskar mRNA, and proteins
such as Y14 and Magoh that bind to this mRNA at the splicing stage appear to function later in determining its cytoplasmic localisation and translation.\textsuperscript{104} Indeed, this is consistent with the finding that several proteins that bind to LSs are hnRNPs or other nucleus–cytoplasm shuttling proteins. Moreover, RNA–protein interactions alter both protein and RNA conformations and so enable further binding and complex formation; this, as well as further protein–protein interactions, is likely to be critical in determining the formation and composition of the localisation RNP complex. As is to be expected, some components of the localisation machinery are conserved between species (eg staufen and ZBP1). A particularly interesting feature of many of the proteins identified as localisation factors (hnRNPs, translation factors) is that they have other functions in post-transcriptional events and thus are multifunctional proteins. To date, it is not clear if localisation occurs largely by different proteins binding to distinct mRNAs or by one protein interacting with a range of mRNAs.\textsuperscript{115}

There appears to be a multiplicity of ways in which the LS exists in the RNA: the signals can be formed by either single regions or multiple elements and both structural elements and repeat sequence motifs are common. In addition, the signals can be comprised of elements in different parts of the 3′UTR, as defined by linear RNA sequence. In certain cases, some of these elements show a degree of functional redundancy. Critically, these cis-acting elements are separate from those that regulate other events such as stability and translation — eg in c-myc and c-fos mRNAs, the determinants of stability and localisation are distinct, whereas, in nanos mRNA, the determinants of localisation are separate from those of translation. Thus, the 3′UTRs should be seen as potentially multifunctional regulatory regions that can contain several distinct cis-acting signals.

Localisation studies are on the brink of the genomic era, bringing in ribonomic approaches. More information on RNA-binding domains in proteins and how mRNA LEs are put together is needed so that bioinformatics and genome information can be used to predict potential localised mRNAs and protein components of the localisation pathway. At present, the apparent lack of consistent motifs making up LEs makes it difficult to use sequence analysis to search for signals in novel 3′UTRs. Detailed information on the precise nature of LSs in a wider range of mRNAs is required so that further available tools (eg UTRdb; Pesole \textit{et al.}\textsuperscript{116}) can be developed to search for similar structures/sequences in databases. Experimentally, identification of key binding proteins in combination with gene array technology should help to identify a range of mRNAs that associate with particular proteins or in specific RNP complexes. Analysis of genome data for LSs and their trans-acting factors provides an opportunity for genomics to contribute to an important aspect of post-transcriptional control of gene expression. Future developments of these approaches will answer the question — ‘how many mRNAs contain localisation signals?’

Acknowledgments
The authors’ work is supported by BBSRC (Grants C13737 and C14236).

References


