Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs

Brenda L. Bass, Susan R. Hurst and Jeff D. Singer

Department of Biochemistry, University of Utah, 50 North Medical Drive, Salt Lake City, Utah 84132, USA.

Background: Although most RNA-binding proteins recognize a complex set of structural motifs in their RNA target, the double-stranded (ds)RNA-binding proteins are limited to interactions with double helices. Recently, it has been discovered that some dsRNA-binding proteins share regions of amino-acid similarity known as dsRNA-binding motifs.

Results: A *Xenopus* ovary cDNA expression library was screened with radiolabeled dsRNA to identify previously uncharacterized dsRNA-binding proteins. The analysis of an incomplete cDNA identified during the screen led to the discovery of two longer cDNAs of related sequence. The proteins encoded by these cDNAs each contained two dsRNA-binding motifs, in addition to an auxiliary domain rich in arginine and glycine. The nucleic-acid-binding properties of a fusion protein containing the two dsRNA-binding motifs and the auxiliary domain were analyzed using a gel mobility shift assay. The fusion protein bound dsRNA of a variety of different sequences, and exhibited a preference for binding to dsRNA and RNA-DNA hybrids over other nucleic acids. Appropriate mRNAs, corresponding to each cDNA, were detected in polyadenylated RNA isolated from *Xenopus* stage VI oocytes, but translation of one of the mRNAs appeared to be masked until meiotic maturation.

Conclusion: dsRNA-binding motifs are often found in proteins that bind dsRNA, and our results show that they can be associated with auxiliary domains rich in arginine and glycine. These motifs can confer very tight binding to dsRNA. Binding can also occur to RNA-DNA hybrids, suggesting recognition of some aspect of the A-form helical structure that is adopted by both dsRNA and RNA-DNA hybrids.

Current Biology 1994, 4:301–314

Background

Many biological processes involve specific interactions between proteins and RNA molecules. RNA-binding proteins usually interact with a complex set of structural motifs in their RNA target (reviewed in [1]). A unique combination of single-stranded loops, bulged nucleotides, internal mismatches and double-stranded stems form the basis of an RNA molecule's three-dimensional structure, and these unique features allow a protein to distinguish it from other RNA molecules.

The double-stranded (ds)RNA-binding proteins differ from most RNA-binding proteins in that, like most DNA-binding proteins, they are limited to interactions with a single structural motif, the right-handed double-helix. Although little is known about the interactions that allow a dsRNA-binding protein to recognize an A-form RNA helix — the helical form adopted by dsRNA — the principles governing the interactions most likely differ from those involved in protein binding to a B-form DNA helix. While the major groove of B-form DNA is quite wide (~12Å; [2]) and freely accessible to a protein, the major groove of A-form RNA is very narrow (~4Å; [3]) and deep, which precludes protein interactions unless the helix is distorted. As most of the functional groups of the bases that allow sequence-specific recognition lie within the major groove [4], the structural differences between A-form and B-form helices suggest that sequence-specific recognition will occur less frequently with dsRNA-binding proteins than with DNA-binding proteins. Indeed, there are now many examples of DNA-binding proteins that interact in a sequence-specific manner with DNA, usually within the major groove [5], whereas none of the dsRNA-binding proteins have shown any obvious sequence specificity.

The dsRNA-binding proteins that have been discovered so far have a wide variety of (sometimes enigmatic) biological roles, but they all interact with dsRNA at some level. For example, activation of the mammalian enzymes dsRNA-dependent protein kinase (PKR) and 2',5'-oligoadenylate synthetase (2',5'-OAS) requires binding of dsRNA to the enzyme both reviewed in [6,7]). The bacterial dsRNA nuclease, RNase III (reviewed in [8]), acts directly on dsRNA to produce single-strand nicks or staggered double-strand breaks. dsRNA is also a substrate for the dsRNA adenosine deaminase (dsRAD), which converts adenosines within dsRNA to inosines; this enzyme is present in organisms throughout the animal kingdom (reviewed in [9]). Certain viruses encode dsRNA-binding proteins, including the reovirus sigma 3 protein [10–12], and the vaccinia virus E3L gene product (p25) [13].

Correspondence to: Brenda Bass.
Through a poorly defined mechanism, viral infection promotes the formation of dsRNA in mammalian cells [14–16], and it is clear that the dsRNA-binding proteins PKR and 25'-OAS are activated by this dsRNA and involved in the host antiviral response (reviewed in [17]). The viral dsRNA-binding proteins p25 [13] and sigma 3 [18,19] are produced by the viruses in an attempt to abrogate this host response by masking the dsRNA. dsRNA-binding proteins are not, however, limited to functions related to viral infection. RNase III is involved in RNA processing and stability, dsRAD seems to be involved in RNA editing [20], and there is much evidence that PKR and 25'-OAS have additional functions in cell growth and proliferation (reviewed in [21]).

Two factors make it difficult to understand the full scope of the in vivo roles of dsRNA-binding proteins. First, the lack of sequence specificity makes it difficult to identify all endogenous RNAs that potentially interact with a dsRNA-binding protein. Although most dsRNA-binding proteins probably require a minimum length of helical dsRNA for binding [22,23], intramolecular as well as intermolecular dsRNAs are potential targets as both bind dsRNA-binding proteins in vivo [6–8]. Second, although short intramolecular duplexes can be identified in many biologically significant RNAs, the source of longer dsRNA in eukaryotic cells is unclear. Some reports suggest mammalian cells contain such RNAs, even in the absence of viral infection [24–26]. Intermolecular RNA duplexes formed between sense and naturally-occurring antisense RNAs exist in bacteria and are clearly substrates for the prokaryotic enzyme RNase III (reviewed in [8]), and naturally-occurring antisense RNA provides a potential way in which long stretches of dsRNA could be formed in mammalian cells (see [20], for example).

With the goal of identifying additional dsRNA-binding proteins, we screened a Xenopus laevis cDNA expression library with radiolabeled dsRNA. This screening protocol led to the identification of two cDNAs with high sequence similarity. Although the encoded proteins do not have extensive sequence similarity to other entries in available databases, they both contain two copies of a recently discovered motif (here and [27]). This dsRNA-binding motif (dsRBM) has been observed in a subset of the dsRNA-binding proteins discussed above and, in addition, has led to the classification of a number of other proteins as dsRNA-binding proteins, including TAR-binding protein (TRBP; [28]) and the Drosophila staufen protein [27]. A unique feature of the Xenopus proteins identified by our screen is the association of the dsRBMs with an auxiliary domain rich in arginine and glycine residues. We have characterized the nucleic-acid-binding properties of a fusion protein that contains the two dsRBMs in association with the arginine-and-glycine-rich domain. In addition, as a first step in understanding the function of the dsRNA-binding proteins in Xenopus, we have monitored expression of the dsRNA-binding proteins during the early stages of Xenopus development.

Results

To identify previously uncharacterized dsRNA-binding proteins, we adapted protocols originally developed for cloning DNA binding proteins [29,30]. These methods rely on the expression of recombinant clones that encode DNA binding proteins in Escherichia coli. Proteins that recognize specific DNA sequences are identified by probing replica filters of the library with the relevant radiolabeled DNA. For our purposes, a Xenopus laevis ovary cDNA expression library was screened with radiolabeled dsRNA. One of the cDNAs identified, cDNA-4F, encodes a protein containing a newly defined dsRNA binding motif and is the subject of this paper.

Two similar oocyte-expressed mRNAs encode proteins with dsRNA-binding motifs

We began our characterization of cDNA-4F by performing a northern analysis of Xenopus laevis ovary polyadenylated (polyA+) RNA. A 32P-labeled oligonucleotide probe, complementary to cDNA-4F, aligned with the predicted transcript size (Fig. 1).

![Fig. 1. Northern blot analyses. A single blot was probed successively with probe a (family-specific), probe b (4F.2-specific) and probe c (4F.1-specific). See Materials and methods for sequences of probes, and Figure 2a for sites of probe complementarity. A*, 10 µg of polyA+ RNA isolated from Xenopus laevis ovaries; i.v., in vitro hybridization control containing 2 ng each of synthetic transcripts corresponding to the sense strands of cDNA-4F.1 and cDNA-4F.2. Positions of size markers are given to the right in kilobases (kb). We assume the 4F.2 in vitro transcript migrates faster than the 4F.2 polyA+ RNA because it was synthesized from an incomplete cDNA. We do not understand the slight mobility difference between the 4F.1 in vitro and polyA+ RNAs, but assume the difference is not due to coding sequences as a protein translated from the in vitro transcript co-migrates with endogenous protein (Fig. 7). Note that the ratio of the two polyA+ transcripts observed with probe a varied in different experiments and may depend on the method of polyA+ selection.](image-url)
hybridized to two transcripts of approximately 3.7 and 2.6 kb (Fig. 1, probe a, lane A'). Identical results were obtained using a probe corresponding to the complete antisense strand of cDNA-4F (data not shown). Neither polyA RNA matched the length of cDNA-4F (1,464 bp), indicating this cDNA was incomplete. To isolate longer cDNAs, the library was rescreened with a radiolabeled antisense 4F transcript using standard nucleic acid hybridization protocols. Two cDNAs, 4F.1 (2,813 bp) and 4F.2 (3,408 bp) were identified. Further northern analyses, using oligonucleotide probes (Fig. 1, probes b and c) specific to either cDNA, demonstrated that cDNA-4F.1 corresponds to the lower molecular weight polyA RNA observed with probe a, and 4F.2 to the higher molecular weight polyA RNA.

The nucleotide sequence of cDNA-4F, as well as that of cDNA-4F.1 and cDNA-4F.2, was determined, and the structures of all cDNAs are shown schematically in Figure 2a. cDNA-4F.1 appears to be a longer version of...
cDNA-4F, and contains an open reading frame (ORF) of 2085 nucleotides, encoding 695 amino acids. We judged the ORF to be complete, as it is preceded and followed by untranslated regions (UTRs) containing multiple myeloma virus (MMV) ORF begins with the first AUG found in any of the three frames, and the surrounding nucleotides (UCAADGUC) exhibit the highly conserved purine at position -3 [51]. The sequences of cDNA-4F that correspond to those in cDNA-4F.1 are not completely identical: after 13 nucleotides of unknown origin, the sequence of cDNA-4F (from nucleotide 1325 to the polyA tail) differs from that of cDNA-4F.1 at two nucleotides. One of these changes (A 1845 to C) produces a conservative amino-acid change, whereas the other (C 2577 to T) is within the 3' UTR. cDNA-4F.1, most likely represents a complete, aligic version of cDNA-4F, but further genomic analyses are required to establish this.

cDNA-4F.2 is a related, incomplete cDNA. An ORF of 2400 nucleotides (800 amino acids) begins immediately at the 5' end of the cDNA, and is followed by a 951-nucleotide-long 3' UTR containing numerous stop codons. The first 1762 nucleotides of cDNA-4F.2 are 93% identical to nucleotides 522-274 of cDNA-4F.1. The sequences diverge significantly near the stop codon of cDNA-4F.1, and the ORF of cDNA-4F.2 is extended at the carboxy terminal.

Based on the length of cDNA-4F.2, and the mobility of the 4F mRNA in the northern-blot analysis, it seems likely that the actual 5' structure of 4F-2 will be similar to that of 4F-1 (dashed outline), thus predicting a full-length cDNA with approximately 521 additional nucleotides at the 5' terminus, encoding an additional 98 amino acids.

The predicted amino acid sequences of 4F.1 and 4F.2 are shown in Figure 2b. Preceding the carboxy-terminal extension of 4F.2, the proteins exhibit 97% similarity across a region of 550 amino acids. The amino-acid sequences show several interesting and significant features. Most importantly, although the full-length cDNA-4F and 4F-2 are not represented in available databases, the amino acids in two regions of each protein (shaded blue in Fig. 2) are similar to each other as well as to a number of other previously characterized proteins. Over the past year, several recent reports have noted these 'motifs' (dsRBMs), and their association with proteins that can bind dsRNA [27,28,32,33]. The two dsRBMs from 4F.1 are shown aligned with other proteins known to contain these motifs (Fig. 3). Recent sequence information for the
Caenorhabditis elegans cosmid K12H4 has revealed an ORF that contains a dsRBM that we have added to the alignment. Interestingly, the C. elegans protein also has additional sequence similarities with RNase III and a number of RNA helicases (see accession number L14331).

A unique feature of the 4F proteins, compared to other proteins containing dsRBMs, is the presence of a domain characterized by a high content of arginine and glycine (RG domain; shaded red in Fig. 2). Although not previously found in dsRNA-binding proteins, similar domains rich in arginine and glycine have been observed in other nucleic-acid-binding proteins, including nucleolin [34], fibrillarin [35,36], SSB1 [37], the GAP-associated tyrosine phosphoprotein p62 [38] and the hnRNPs A1[39] and U [40]. The 4F proteins also contain a bipartite nuclear-localization sequence (reviewed in [41]), and within the carboxy-terminal extension of 4F-2 is a stretch of 69 amino acids characterized by a high content of prolines (25%) and glutamines (33%).

Analysis of dsRNA binding by gel mobility shift assays

We used the protein encoded by the truncated cDNA, cDNA-4F, to study the nucleic-acid-binding properties of the 4F proteins. As this cDNA was identified in our initial screening, we reasoned that the protein it encodes must contain all of the amino-acid determinants necessary for binding to dsRNA. Furthermore, in addition to the two dsRBMs, this protein contains the associated RG domain. Previous studies of RG domains show they can bind single-stranded as well as double-stranded RNA and DNA, with little specificity [42], although in some cases a slight preference for single-strands and RNA is observed [43]. Among the proteins known to contain RG domains are those in which the domain appears to be entirely responsible for the nucleic-acid-binding properties of the protein [40], and others in which the domain is associated with other RNA-binding motifs, such as the RNA recognition motif (RRM) [39,43] or a putative zinc-binding domain [44]. In the latter examples, the affinity of the complete protein appears to be the sum of the affinity of both types of RNA-binding motif. Thus, we considered it important to include this domain in the protein used for our binding studies.

To obtain large quantities of the protein, cDNA-4F was subcloned into the vector pMAL-c2 (New England Biolabs), designed for over-expression of proteins fused to maltose-binding protein. The fusion protein (MBP-4F), a 723 amino-acid protein (79023 KD) containing all 335 amino acids encoded by cDNA-4F, was used in gel mobility shift assays. A synthetic 32P-labeled 36 bp dsRNA (Fig. 4a) or 102 bp dsRNA (Fig. 4b) was mixed with increasing amounts of MBP-4F, and subsequently electrophoresed on a native 4% polyacrylamide gel. The 36 bp and 102 bp dsRNAs were unrelated, with no internal repetition, and the sequences did not correspond to any known biologically-relevant sequence. For both dsRNAs, an RNA-protein complex

![Fig. 4. Mobility shift analyses of MBP-4F binding to dsRNA. (a) A 36 bp dsRNA (10 pM) was mixed with varying concentrations of fusion protein as indicated. The electrophoretic mobilities of the free and complexed dsRNA are shown. The smear between the two complexes may indicate some dissociation of the lower affinity complex during electrophoresis [63]; see Materials and Methods. (b) As in (a) except the dsRNA was 102 bp. (c) Primary data for binding to the 36 bp dsRNA, similar to the data in (a), were generated in multiple experiments and quantified. Each point is an average of four determinations, except the three highest protein concentrations, which are an average of three, two and two determinations (from lowest to highest). Error bars are +/- standard deviations. Note that MBP-4F was only 30% pure, and densitometry was used to estimate the molar concentration of full-length fusion protein in various preparations (see Materials and methods). We considered this estimate to be suitable for these binding studies, as western analyses indicated the contaminating species were degradation products of MBP-4F, and other analyses indicated all observed shifts resulted from the intact fusion protein (data not shown). All intact MBP-4F was assumed to be competent for dsRNA binding.](image-url)
was first observed at -0.1 μM protein. As the protein concentration was increased, additional complexes were observed, with a total of two complexes observed for the 36bp dsRNA and five complexes for the 102bp dsRNA. The number of complexes observed with the different RNAs suggested the minimum binding site is 18-20bp long, with multiple proteins able to occupy a single RNA molecule (two for 36bp, five for 102bp). An alternative explanation, that the multiple shifts were due to a single protein binding multiple RNA molecules, was not consistent with the fact that all of the 36bp dsRNA had bound protein (leaving no free RNA) before the second complex could be detected.

Several experiments similar to that for which the results are shown in Figure 4a were repeated to determine the dependence of binding on protein concentration (Fig. 4c). The data were used to calculate an apparent dissociation constant (Kd) of 2.7 x 10⁻⁹ M for the interaction corresponding to the first shift observed with the 36bp dsRNA. Thus, MBP-4F can bind tightly to dsRNA. We note that binding of the 4F polypeptide to dsRNA was not dependent on the presence of the MBP sequences (see Fig. 6b). Further, when the MBP domain was removed using factor Xa, RNA-protein complexes of increased electrophoretic mobility were observed, with little change in the affinity for dsRNA (data not shown).

The protein encoded by cDNA-4F prefers to bind dsRNA and RNA–DNA hybrids

Enzymatic activity assays indicate that previously characterized dsRNA-binding proteins, although they lack sequence specificity, are quite specific for dsRNA. For example, the kinase activity of PKR cannot be activated by dsRNA [45-47]. RNA–DNA hybrids [45,46] or single-stranded RNA [45,46] (except those with regions of intramolecular dsRNA [48]). Assays of enzymatic activity suggest a similar specificity for 25S-OAS [49,50] and dsRAD [51,52].

Our gel mobility shift assay allowed us to study binding specificity of MBP-4F directly, rather than indirectly by enzymatic activity. As expected, we found no evidence of sequence specificity and observed that MBP-4F could shift dsRNA of a variety of different sequences (see, for example, Figures 4a, 4b and 6b). However, as demonstrated in competition studies (Fig. 5a), MBP-4F exhibited a strong preference for binding to dsDNA and RNA–DNA hybrids. Two series of 3H-labeled nucleic acids were tested for their ability to compete with the 3P-labeled 36bp dsRNA in a mobility shift assay. One series consisted of molecules of ~100 bases (or bp) in length, while the other series was ~800 bases (or bp) in length.

To eliminate the possibility of the relative binding being affected by the nucleic acid sequences, each series of

---

**Fig. 5.** Competition studies. (a) Mobility shift assays as presented in Fig. 4, except the concentration of 32P-labeled 36bp dsRNA was 50μM. 3H competitors were added at 1X or 10X molar excess as indicated, and fusion protein was constant at 0.3μM. 32P and 3H nucleic acids were mixed prior to the addition of MBP-4F. Lanes corresponding to the addition of different 100mer or 600mer nucleic-acid competitors are indicated: --, no competitor; RR, dsRNA; DD, dsDNA; R, single-stranded RNA; RD, RNA–DNA hybrid; ΔR, heated single-stranded RNA. (b) Similar to (a) except fusion protein (0.25μM) was incubated with a mixture of 32P-labeled 36bp dsRNA and 32P-labeled 100mer dsRNA, or 32P-labeled 36bp dsRNA and 32P-labeled 100mer RNA–DNA hybrids. The 36bp dsRNA was 50μM and the 100mer competitors were at the concentrations (μM) indicated. The electrophoretic mobility of the free nucleic acid species (RR, RD) and the nucleic acid–protein complexes (RR·P) are indicated. The 100RR·P complex has a slower mobility than the 36RR·P complex because of the different lengths of the dsRNA in the two complexes.
molecules was related in sequence. Thus, the ability of a dsRNA to compete was compared with that of the dsDNA fragment encoding it, with one of its RNA strands and with an RNA–DNA hybrid formed between one of its strands and the complementary DNA strand. With both competitor series, at 1x and 10x molar excess of competitor to 36pb dsRNA, the mobility shift was almost completely competed by dsRNA and RNA–DNA hybrids. Very little competition was observed with a 1x molar excess of the 100mer dsDNA and ssRNA, but these molecules began to compete at higher molar ratios or when the length of the competitor was increased (from 100mer to 800mer). For both series of competitors, dsRNA and RNA–DNA hybrids were the most effective competitors, followed by ssRNA and finally dsDNA. This trend was reproducible in multiple experiments (data not shown).

We considered the possibility that the low-affinity competition with a single strand of RNA was due to the existence of intramolecular double-stranded regions. To explore this possibility the single-stranded competitors were, in some cases, heated for 90 seconds at 90°C and rapidly cooled on ice prior to adding to the reaction mixture (ΔR). No difference was observed between the heated and unheated molecules in their ability to compete. However, this result was inconclusive as internal structure could reform during the 20 minute incubation, and at present the nature of the single-strand competition is unclear.

Because of the large amount of competition by dsRNA and RNA–DNA hybrids in the experiment illustrated in Fig. 5a, it was hard to determine if MBP-4F had different affinities for these molecules. Thus, further competition assays were performed at lower ratios of 100mer competitors to 36pb dsRNA (Fig. 5b). In this experiment, the 36pb dsRNA substrate, as well as the 100mer dsRNA and RNA–DNA competitors, were all labeled with 32P to allow careful quantification of their relative amounts. Although the 100mer dsRNA consistently competed better than the 100mer RNA–DNA hybrid, the difference was slight: averaging three similar experiments showed, at 10μM competitor and 50μM 36pb dsRNA, the 36pb dsRNA–protein complex was reduced 57% by 100mer dsRNA and 36% by the 100mer RNA–DNA hybrid. Surprisingly, as illustrated in Fig. 5b, we observed that although the RNA–DNA hybrid was an effective competitor, it did not produce a mobility shift. This observation held true even at higher concentrations of protein (440μM, data not shown).

In summary, the competition studies indicate the 4F proteins bind tightly to dsRNA and RNA–DNA hybrids, but poorly to ssRNA and dsDNA. As the helical structure of RNA–DNA hybrids is thought to be similar to that of A-form dsRNA (see Discussion), these studies suggest some aspect of the A-form helix is required for binding.

Regions of protein 4F.1 necessary for binding to dsRNA
Although the 4F sequence in MBP-4F differed at one residue from the corresponding sequence of 4F.1, the substitution was conservative (M to I; see Fig. 2b) and resulted in identity to 4F.2. Thus, given the results shown in Figures 4 and 5 it seemed likely that the carboxy-terminal half of the protein encoded by the complete cDNA (4F.1) could bind dsRNA. However, we wished to verify that dsRNA-binding could occur in the context of the complete protein, to find out if other regions of the protein contributed to the interaction with dsRNA and whether the dsRNA-binding motifs could bind dsRNA in the absence of the auxiliary RG
domain. To answer these questions, another mobility shift assay was performed. This assay differed from those performed previously in that the protein, rather than the nucleic acid, was radiolabeled, and the protein was not an overexpressed fusion protein but was synthesized in vitro using reticulocyte lysates.

mRNAs corresponding to the complete cDNA and various truncated cDNAs were synthesized in vitro and used to generate the proteins shown in Figure 6a. The electrophoretic mobility on a native polyacrylamide gel (Fig. 6b) of each 35S-labeled protein was assayed after incubation alone (-), or in the presence of the 800mer dsDNA (D) or the 800mer dsDNA (D). Although the proteins were poorly resolved on the native gel, clear nucleic-acid-dependent changes in mobility were observed. For example, the mobility of the full-length protein (4f1) was altered by dsRNA but not dsDNA, establishing that dsRNA binding could occur within the context of the entire 4f1 protein.

In contrast, no change in electrophoretic mobility was observed when either dsRNA or dsDNA was added to the amino-terminal half of the 4f1 protein (N; the portion of the protein not included in MBP-4f). The latter result suggested that all determinants for dsRNA-binding were within the boundaries encoded by cDNA-4f. As expected, the region of protein 4F included in MBP-4f (3D) also bound dsRNA when assayed as an in vitro translation product; this result emphasizes that the MBP sequences are not necessary for MBP-4f to bind dsRNA. Further truncation of the protein to delete the RG domain (2D) produced a protein that also bound dsRNA, suggesting the dsRBMs are the major determinants of binding. However, when the protein was further truncated to include only the amino-terminal dsRNA binding motif (1D), binding was substantially reduced.

Protein expression during early Xenopus development

As a first step towards understanding the biological role of the proteins encoded by cDNA-4f and cDNA-4f, a polyclonal antibody was raised against MBP-4f. The antibody was used in a western analysis of proteins extracted from nuclei or cytoplasm of Xenopus stage VI oocytes (Fig. 7a). The antibody recognized a single polyepitide in stage VI oocytes that was entirely localized to the oocyte nucleus. The immunoreactive protein had a slower electrophoretic mobility than a protein translated in vitro from 4f1 mRNA, but had a mobility consistent with the predicted molecular weight of 4f2 (assuming aberrant mobility; see Fig. 6a).

Changes in translation of maternal mRNAs sometimes occur during meiotic maturation, the process in which an oocyte becomes an egg. Thus, further western analyses were performed using proteins extracted from eggs, and embryos at various stages of early development. As shown in Figure 7b, in addition to the putative 4f2 protein, a protein that comigrated with the in vitro translated 4f1 was evident in the egg and persisted throughout early embryogenesis. It should be noted that, although the northern analysis of Figure 1 was performed with poly A RNA from total oocytes, both mRNAs were also observed using poly A RNA from stage VI oocytes (data not shown).

Discussion

We have described the identification of cDNAs encoding two new members of a family of proteins characterized by motifs that allow binding to dsRNA. Each of the encoded proteins contain two dsRBMs and an auxiliary domain rich in arginine and glycine. Using gel mobility shift assays, we have characterized the nucleic-acid-binding properties of a fusion protein containing both dsRBMs and the associated RG domain. The fusion protein binds tightly to dsRNA, exhibiting an apparent Kd of ~0.3 nM for a synthetic 36bp dsRNA. Although the fusion protein does not show any apparent sequence specificity, competition studies suggest a preference for binding to A-form helices. The RG domain is not required for dsRNA binding, and although good binding is observed with two dsRBMs, binding is reduced with the amino-terminal motif alone.

Nucleic-acid-binding properties of proteins containing dsRBMs

Our mobility shift analyses allow certain predictions to be made about how the 4f dsRNA-binding proteins interact with dsRNA. First, the observation that two
RNA–protein complexes form with a 36bp dsRNA, and five with a 102bp dsRNA, suggests a minimal binding site of 18–20bp per protein. Of course, at present we cannot rule out the possibility that the minimal binding site is slightly larger, with the final binding event occurring by low affinity binding to a sub-optimal number of nucleotides, or to a site formed by the juxta-position of the ends of two separate molecules. Second, the protein concentrations required to produce the various complexes observed with the 36bp and 102bp dsRNAs (Fig. 4a and b) suggest binding by the 4F proteins may be similar to that observed with non-sequence-specific DNA-binding proteins, and described using conditional probabilities [55].

According to this hypothesis, initial binding would be random along the entire length of the dsRNA, and the protein concentration required for occupation of subsequent sites would be dictated by the probability that the minimal length of nucleotides required for binding was available. This probability will increase with the length of the dsRNA, and consistent with this idea, a second binding event occurs much more readily with the 102bp dsRNA than with the 36bp dsRNA. The last shift observed with both dsRNAs occurs at approximately 100 nM protein, suggesting the conditions governing binding to this site are similar, perhaps requiring sliding of previously-bound protein to reveal a single, minimal binding site. The intermediate shifts observed with the 102bp dsRNA do not show any obvious cooperativity, but further experiments are necessary to confirm this.

A perplexing observation is the almost complete disappearance of the 36bp dsRNA–protein complex on the addition of an equal molar amount of the 102bp dsRNA as competitor (Fig. 5a and b). In these experiments the protein was present in 5–6-fold molar excess of the radiolabeled 36bp dsRNA, and only a single gel shift was observed with each dsRNA, suggesting multiple binding events were not occurring (Fig. 5b). We have not determined a Kd for binding of MBP-4F to the 102bp dsRNA. However, even assuming the 102bp dsRNA could titrate an equal molar amount of protein, the large amount of competition observed in Figure 5a and b would not occur unless the protein was near saturation. Although we do not know if all the fusion protein is active, other experiments have shown that 50 nM 36bp dsRNA is not saturating at protein concentrations near the Kd (data not shown). We are currently analyzing the results shown in Figure 5 more carefully, and exploring the possibility that, as with the dsRNA-binding protein PKR [54], productive binding to MBP-4F is inhibited by high concentrations of dsRNA.

The observation that MBP-4F binds both dsRNA and RNA–DNA hybrids is of particular interest. Although controversial, it is generally accepted that RNA–DNA hybrids adopt conformations closer to A-form than B-form. For example, recent studies indicate the RNA strand in a hybrid adopts an A-form conformation, whereas the DNA strand exhibits sugar puckers intermediate between those characteristic of A-form and B-form helices [55]. Our studies indicate that, under our experimental conditions at least, the structures of dsRNA and RNA–DNA hybrids must be similar at some level as they are both recognized by MBP-4F. At present it is unclear what features of dsRNA and RNA–DNA hybrids allow binding in preference to ssRNA and dsDNA. For example, it is not clear if binding requires only an A-form helical geometry, or if specific features of RNA, such as the 2'-hydroxyl, are required.

Clearly the interactions involved in MBP-4F binding to dsRNA and RNA–DNA hybrids are not identical, as RNA–DNA hybrids can compete for dsRNA binding but cannot themselves form complexes that are stable under our electrophoretic conditions. Recent studies of Escherichia coli RNase H [56] show that this protein binds its natural substrate, an RNA–DNA hybrid, as well as dsDNA and dsRNA, but has a different exchange rate when binding to the non-physiological substrates. Although not yet tested, differences in exchange rate could also explain the differential binding of MBP-4F to dsRNA and RNA–DNA hybrids. As mentioned, previous studies of PKR, which contains two dsRBMs, show that RNA–DNA hybrids cannot activate PKR kinase activity. In the light of our results, we suggest RNA–DNA hybrids may indeed bind to PKR, but not in a functionally competent manner.

Most of our assays used a fusion protein containing two dsRBMs as well as the associated RG domain. However, we have not noticed significant differences in binding affinity with proteins lacking the RG domain (Fig 6b and data not shown). Thus, the 4F proteins may differ from other proteins that contain both RNA-binding motifs and auxiliary RG domains, the affinities for RNA of which are dependent on both types of motif [39,43,44]. When associated with the RRM, the RG domain has also been proposed to lend cooperativity to the overall nucleic acid binding [39,43]; at present we cannot rule out the possibility that the RG domain lends cooperativity to the 4F proteins.

As the binding of MBP-4F to dsRNA appears to be dictated by the dsRBMs, it is of interest to compare its binding properties to those of other proteins that contain dsRBMs. With one exception [28], previous binding studies of dsRBMs have not been performed by mobility shift analyses. Rather, northwestern or chromatographic assays have been used. Although these types of assays did not allow quantification of binding affinities, or resolution of possibly heterogeneous mixtures of RNA–protein complexes, they showed dsRNA-binding is intrinsic to dsRBMs, and suggested not all motifs are equal.

Thus, for example, single motifs of the dsRNA-binding proteins staufen and Xlphb can bind dsRNA, requiring the ~65 amino acids that comprise the entire dsRBM [27]. In contrast, optimal binding of dsRNA by PKR requires both of the two motifs in this protein, but
mutations in the most amino-terminal motif are more debilitating than those in the second motif [33]. Surprisingly, a synthetic protein of 24 amino acids containing only the carboxy-terminal domain of one of the motifs of TRBP is sufficient to mimic the binding observed with the intact 345 amino acid TRBP [28].

Although we have not yet studied the binding of the second dsRBM of the 4F proteins in isolation, we have demonstrated that the amino-terminal motif alone binds dsRNA, albeit poorly (Fig. 6b). Taken together, these studies suggest different motifs differ in their dsRNA-binding abilities.

Based on our studies of MBP-4F, we would predict that other proteins containing dsRBMs will exhibit a preference for binding to A-form helices. However, in light of the studies discussed above, we speculate that not all proteins containing dsRBMs may bind as tightly to dsRNA as MBP-4F. In only one other case has the affinity of a dsRBM protein been quantitatively determined [54], and here a different assay and an RNA of heterogeneous length were used, making comparisons difficult. In determining the Kd for binding of MBP-4F to the 36bp dsRNA, we made several assumptions in regard to protein concentration (see Fig. 4 legend); thus future studies may allow refinement of the calculated dissociation constant. Regardless, our dissociation constant represents a value with which to compare future studies on 4F.1 and 4F.2, as well as other proteins that contain dsRBMs.

Biological significance of proteins 4F.1 and 4F.2

At present we do not know the physiological role of the 4F.1 and 4F.2 proteins in Xenopus laevis eggs. Both proteins contain nuclear-localization sequences, suggesting they play a role in the nucleus. Although we have not determined if both proteins are always nuclear, an antibody raised against MBP-4F recognizes a single protein that is exclusively localized to the nucleus in stage VI oocytes. Our studies indicate the 4F group of proteins can bind to as little as 18-20 bp. This is a bit longer than most stems that occur in biologically-relevant RNAs, but we have not established if binding can occur to shorter lengths of dsRNA. Thus, intramolecular dsRNA regions, in addition to the longer dsRNA regions that can form between sense and antisense RNAs, are potential substrates of the 4F proteins.

Although both 4F.1 and 4F.2 mRNAs exist in the oocyte, western analyses of proteins from these cells identify only a single polypeptide; two immuno-reactive proteins are evident in eggs and early embryos. Although further studies will be required to understand these results, the gel mobilities of the immunoreactive bands suggest that 4F.2 is the protein found in oocytes, eggs and embryos, and 4F.1 is the protein lacking in oocytes.

As in many organisms, transcription does not occur during early Xenopus embryogenesis, and the embryo must rely on maternal mRNAs stockpiled in the oocyte.

One means of regulation during early development is the translational silencing (masking) or activation (unmasking) of maternal mRNAs (see reviews in [57]. Although we cannot rule out mechanisms that act subsequently to translation (such as protein lysis), the regulation of the 4F.1 protein level is consistent with trends observed for masked messages. For example, changes in translation of masked messages often occur during meiotic maturation, as the stage VI oocyte becomes an egg. In fact, the pattern of 4F.1 protein expression appears similar to that of Xenopus cyclin A [58]. For a number of masked messages, translational control involves cis-acting sequences in 3' UTRs. Although the 4F.1 and 4F.2 mRNAs are very similar in sequence in the coding regions, their 3' UTR sequences are very different, allowing for the possibility that these sequences mediate differential translation.

We have considered the possibility that the cDNAs we have cloned encode previously characterized dsRNA-binding proteins. Xenopus dsRAD has not yet been cloned, but the protein has recently been purified [59]. The MBP-4F antibody does not cross-react with purified dsRAD, and partial amino-acid sequence analysis of dsRAD reveals no homology to the 4F proteins (R. Hough and B.L.B., unpublished data). We also note that the proteins described here differ in sequence from another, recently described Xenopus dsRNA-binding protein, Xlps2 [27].

We do not yet understand the organization of the genes encoding dsRNA-binding proteins 4F.1 and 4F.2, or the functional relationship of these two very similar proteins. The sequence divergence between 4F.1 and 4F.2 cDNAs suggests that they are not derived from alleles, but rather from two distinct genes that arose by duplication, a common phenomenon in the pseudo-tetraploid genome of Xenopus laevis [60]. Assuming that the 4F.1 and 4F.2 proteins are encoded by different genes, it is possible that the nascent transcript of each gene is subject to alternative processing to give two mature RNA species; if so, the two cDNAs described here, 4F.1 and 4F.2, would each represent one of the two spliced products from each of the different genes.

Conclusions

An increasing number of dsRNA-binding proteins are being discovered, and their sequences found to contain dsRBMs. Our results show that, like other RNA binding motifs, dsRBMs can be associated with arginine and glycine rich sequences. Although the nucleic-acid-binding properties of all dsRBMs proteins have not yet been studied, the studies described here suggest the proteins as a group will exhibit a preference for binding A-form helices, and in some cases, exhibit very high binding affinities for dsRNA. The cloning of two previously uncharacterized proteins that contain dsRBMs will facilitate the understanding of the various biological functions of this newly defined group of proteins.
Materials and methods

Cloning and sequencing of cDNAs

A Xenopus laevis ovary cDNA expression library, constructed in Agt11 SfiI-Nor (Promega), was provided by Richard Harland. The phage containing cDNA-4F was identified by screening the library with a radiolabeled dsRNA probe (~800bp, see Nucleic acid preparation). Protocols were essentially those developed for screening expression libraries with radiolabeled DNA (native conditions; [61]), with modifications to guard against ribonucleases. The 10x binding buffer was replaced by a buffer containing at 1x 10mM Tris (pH 7.5), 1mM EDTA, 50mM NaCl, 0.04% BSA, 0.04% Ficoll 400 and 0.04% polyvinyl pyrrolidone; non-fatt milk was excluded during all procedures. For all steps requiring the addition of competitor salmon sperm DNA, torula RNA was also added at a concentration of 10ng ml⁻¹; both competitors were treated to remove containing ribonucleases. Washing steps were performed with the 1x buffer described above.

The cDNA clones 4F.1 and 4F.2 were cloned by rescreeing the library using nucleic-acid hybridization conditions. Plating and filter preparation were as described [62]. Filters were incubated at 40°C for 2.5-3 hours in prehybridization solution (25mM KH₂PO₄, at pH 7.4, 5x SSC, 5x Denhardt, 50 µg ml⁻¹ each of salmon sperm DNA and torula RNA, 30% diconitrated formamide, 0.1% SDS), then at 40°C for 2.5-3 hours with hybridization solution (prehybridization solution plus 10% dexam sulfate) that contained a radiolabeled RNA corresponding to the antisense sequence of cDNA-4F (see Nucleic acid preparation). Filters were washed twice for 15 minutes at room temperature in 1x blot wash (10x consists of: 0.5M Tris (pH7.8), 0.02M EDTA (disodium), 5% sodium pyrophosphate, 0.5% Sarskol, 0.5% SDS, 0.2% Ficoll, 0.2% polyvinylpyrrolidone).

For each phage clone (4F.1, 4F.2, the SfiI-Nor fragment containing the cDNA was subcloned into the corre- sponding site of pGEM 13Zf(+) (Promega) to facilitate DNA and RNA preparation. Both strands of each cDNA were sequenced using Sequenase 2.0 kits (USB). Initial sequence was obtained using pUC/M13 forward and reverse primers (Promega) complementary to pGEM-13Zf(+) sequences flanking the cDNA inserts. Subsequent primers were selected every 300-350bp on both strands.

Northern analyses

Ovaries excised from X. laevis females that had been injected with 50 units pregnant mare serum gonadotropin 18 hours earlier were incubated overnight at room temperature in MBS (10mM Hepes at pH7.5, 88mM NaCl, 1mM KCl, 2.4mM NaHCO₃, 0.41mM CaCl₂, 50 µg ml⁻¹ Gentamicin sulfate) containing 1 mg ml⁻¹ collagenase B. Total RNA was isolated using the protocol described [61]. PolyA⁺ RNA was selected using an Oligotexdt ktl (Qiagen).

Electrophoresis and northern blot transfer were performed according to standard protocols [62]. Prehybridization and hybridization were as described for cloning 4F.1 and 4F.2 except that incubations took place at 42°C and probes were end-labeled DNA oligonucleotides. The sequences of the oligonucleotide probes used are given below, 5' to 3', with numbers indicating complementary sequences corresponding to those shown in Fig. 2a.

(a) Family-specific:
TCTTGCCTGGAGCCTCTGCTGATCGTGT (1995-2025)
(b) 4F.2-specific:
AGTCGCTCTCCAGACAGATGTGCTCTCT (5226-5255)
(c) 4F.1-specific:
AGTGATCCTGAAGACAGTGGTCTCCTG (5283-5211)

Oligonucleotide probes were labeled at their 5' termini using T4 polynucleotide kinase and [γ-³²P] ATP, then passed over a Chroma spin-10 column (Clontech) to remove unincorporated nucleotides. All probes were completely complementary to target RNAs, except the family-specific probe which had two mismatches when hybridized with 4F.2. After hybridization, the membranes were washed at 50°C in 6x SSC, pH7.0.

Preparation of MBP-fusion protein

pGem-4F was digested with Ncol, and partially with EcoRI. A Ncol-EcoRI fragment (1477bp) encompassing the entire cDNA was filled-in using the Klenow fragment of E. coli DNA polymerase I, gel purified and ligated into the Xmal site of the pMAL-c2 expression vector (New England Biolabs). The soluble fusion protein was expressed and purified using amylose affinity chromatography as described [62]. The fusion protein was dialyzed against 10mM Tris (pH 8.0), 100mM NaCl, 1mM DTT. Samples were stored at -20°C after addition of glycerol to 20% Total protein concentration was determined at A280 using the theoretical extinction coefficient of 1A280=0.095mg ml⁻¹. From densitom- etry of SDS-PAGE gels we estimated that intact fusion protein represented ~80% of our protein preparations. This value was used to determine the molar concentration of the fusion protein in various preparations.

Preparation of in vitro translated proteins

mRNAs for translation were transcribed from linearized pGem-4F and pGem-4F.1 using SP6 polymerase (see Nucleic acid preparation). pGem-4F.1 was digested with Ncol and NotI to generate templates for 4F.1 and N, respectively. pGem-4F was digested with Ncol, NotI, and BsrI to generate templates for 4F.1, 2D, and 1D, respectively. As pGem-4F contained these BsrI sites and two NotI sites, these digestion products were gel purified to ensure transcription of only the correct DNA. In vitro translations were done in nucleate treated Rabbit Reticuloocyte Lysate systems (Promega) using 35S- methionine as label. Translated protein concentrations were calculated by TCA precipitation. The radioactivity of proteins resolved by SDS-PAGE was quantitated to estimate relative protein concentrations.

Mobility shift assays

Mobility shift assays with MBP-4F were performed in 10µl of a Mobility Shift Buffer (MSB) containing a final concentration of 10mM Tris (pH 8.0), 25mM KC1, 10mM NaCl, 1mM MgCl₂, 0.5mM DTT, 16mM glycerol, 100µg ml⁻¹ BSA, 200µM ATP (magnesium and ATP do not appear to be essential for mobility shifts). Fusion protein and nuclear acids were added at concentrations given in the figure legends. When necessary, the fusion protein was diluted into MSB prior to addition to the binding reaction. In preliminary experiments, a time course of incubation times demonstrated that no change in fraction bound occurred during a 1-60 minute time course at 25°C; thus, for convenience, binding was performed at 23-25°C for 20 minutes. Binding reactions were stopped by loading directly onto a 4% polyacrylamide native gel. Dyes were excluded from samples, as initial
experiments indicated they were inhibitory to binding. Electrophoresis was for 2 hours at 150 volts on a gel of 18 cm in length; running buffer was 1x TBE (10X is: 1 M Tris Base, 0.83 M H₃BO₃, 0.01 M EDTA). Gels were dried and then subject to autoradiography.

Mobility shift assays using \textit{in vitro} translated proteins were performed similarly in MSB. Proteins were added as aliquots of the \textit{in vitro} translation reactions to give equal amounts of the proteins being compared. The amount of \textit{in vitro} translation reaction components added was equalized by adding \textit{in vitro} translation mix that had not been programmed with mRNA. Binding reactions were incubated and processed as described for fusion-protein binding reactions, except electrophoresis was for 4 hours at 200 volts, and after electrophoresis, gels were soaked in H₂O₂ to remove unincorporated [³H]Sminethionine.

Radioactivity in dried gels was quantitated using a Molecular Dynamics PhosphorImager. The fraction bound was determined either by dividing the radioactivity in the dsRNA-free band by the radioactivity in the entire lane (fraction bound) or by dividing by the amount of radioactivity in a region encompassing the RNA complex by this amount plus the dsRNA free radioactivity (fraction bound = complex/complex + RNA in lane). Both methods gave identical Kd values, indicating the equilibrium was not disturbed by our electrophoretic conditions. To determine Kd values, a SigmaPlot (Jandel Scientific) curve-fitting program was used to fit the data to the equation: fraction bound = 1/(1 + (Kd/protein)p), where m = 1 for monomer binding. The concentration of dsRNA was kept constant at 10μM, and it was assumed that protein bound was approximately equal to protein free; thus the protein concentration required for 50% binding was assumed to be very close to the Kd (pp) [65].

Nucleic acid preparation

\textbf{DNAmer:} The 36 bp dsRNA consisted of the following hybridized strands:

5′ GAA TAA ACA GAG AAA AAA ACA AGG ACA GGG ACC AGG
UUU AUA GUG UGC UUU UUC UUC UUC UUC GGC GGC GGU GGU

Each strand was synthesized from a partially single-stranded DNA template essentially as described [64]. Single-stranded RNAs were gel purified, then hybridized by mixing approximately equal molar amounts in 10mM Tris (pH 7.5), 0.5mM EDTA, followed by heating at 55°C for 10 minutes and slow cooling to room temperature. The resulting 36 bp dsRNA was gel purified, quantitated, then treated with calf intestinal phosphatase, followed by 5′ end-labeling as described for oligonucleotide probes.

\textbf{800mer series:} An RNA duplex of approximately 800bp, corresponding to the hybridized sense and antisense transcripts of the chloramphenicol acetyltransferase (CAT) gene, was prepared and gel purified as described [52], except the RNAs were not capped. CAT dsRNA was used as a probe for expression library screening was labeled by including 250μCi each of [³²P]ATP and [³²P]UTP in the transcription reaction for each strand. CAT dsRNA used in competition studies and mobility shift assays (Fig. 6b) was radiolabeled similarly, using 100μCi of [³²P]UTP for each strand.

The CAT sense RNA transcribed as above [³²P]UTP was used as the ~800bp single-strand competitor. 800mer dsRNA was prepared by gel purifying the \textit{SacI}–\textit{BamHI} fragment of pSP55CAT A [65], which encompasses the CAT gene. An ~800bp RNA–DNA hybrid corresponding to the CAT sequences was prepared by mixing the \textit{SacI}–\textit{BamHI} DNA fragment with an excess of the CAT sense transcript in Formamide Binding Buffer (PBB; 80% denatured formamide, 40mM PIPES at pH 6.7, 400mM NaCl, 1mM EDTA). The mixture was heated at 65°C for 5′ then placed at 45°C overnight. Electrophoretic separation showed a band of RNA–DNA hybrid to sense RNA (dsDNA) which was gel purified and confirmed to be an RNA–DNA hybrid by RNAase H sensitivity.

\textbf{100mer series:} A phlluisci vector II KS vector (Stratagene) containing a short insert (of no biological significance) between the T7 and T3 promoters was provided by Janice Pata and was used to prepare the 100mer substrates and competitors. A \textit{BglII} fragment (139 bp) that included both T7 and T3 promoters and intervening template sequences was gel purified and used as a DNA competitor and as a template for 100mer T7 and T3 transcripts. For transcription, the \textit{BglII} fragment was incubated with 2μm ³²P of either T7 or T3 RNA polymerase according to manufacturer's specifications (Stratagene). 50μCi of [³²P]UTP was included in each transcription reaction. Transcripts were used as single-strand competitors (T3 transcript) or hybridized in FIB as above, to form the T7–T3 dsRNA of 102bp, and 5′ overhangs as shown.

\textbf{Western analyses:} Oocytes were isolated from ovaries as above (Northern analyses) and eggs and embryos as described [52], except frogs were primed by injecting pregnant mare gonadotropin (50 units), followed approximately 3 days later by a booster injection of human chorionic gonadotropin (750 units) to induce egg laying. Unless indicated, proteins for western analyses were prepared by placing ten oocytes, eggs or embryos into 16% glycerol, 5% 2-mercaptoethanol, 2.3%
Acknowledgments. We are grateful to R. Harland for providing the Xenopus ovary library, J. Pati for providing plasmid JP11-1, and L. Saccomanno for providing dissected Xenopus nuclear and cytoplasmic fractions. We thank B. Graves, W. Sundquist and the members of our laboratory for their critiques of this manuscript, T. Fermosa and D. Gottschling for technical advice, and J. Cherry, T. Doak, G. Herrick, M. Krause, and H. Swerdlov for helpful discussions. This work was supported by funds to B. L. B. from the National Institute of Health (N.I.H. GM 44073), the Pew Scholars Program and the David and Lucile Packard Foundation. DNA synthesis was performed by the Utah Cancer Center Protein/DNA core facility, supported by the National Cancer Institute (5PO1 CA02101).

References


