J. Mol. Biol. (1997) **271**, 195–208

JMB

A *Xenopus* Zinc Finger Protein that Specifically Binds dsRNA and RNA-DNA Hybrids

Patrick J. Finerty, Jr and Brenda L. Bass*

Department of Biochemistry and Howard Hughes Medical Institute, 6110a EIHG Building 533, University of Utah, Salt Lake City, UT 84112, USA

Proteins containing C_2H_2 type zinc finger motifs represent one of the largest classes of nucleic acid-binding proteins found in nature. We describe a novel zinc finger protein, dsRBP-ZFa, isolated by screening an expression library with dsRNA. The dsRBP-ZFa cDNA encodes a protein containing seven zinc finger motifs and an acidic C-terminal domain. Mobility shift experiments demonstrate that dsRBP-ZFa binds dsRNA and RNA-DNA hybrids with nanomolar dissociation constants and in a sequence independent manner. We also show that DNA and single stranded RNA fail to compete with dsRNA for binding suggesting dsRBP-ZFa prefers to bind an *A*-form helix. Using western analyses we have localized dsRBP-ZFa primarily to the nucleus of *Xenopus laevis* oocytes. The identification of dsRBP-ZFa provides the first example of a zinc finger protein that is specific for dsRNA. In addition, dsRBP-ZFa does not contain the previously described dsRNA binding motif, suggesting certain zinc fingers may provide an alternative way to recognize the *A*-form helix.

© 1997 Academic Press Limited

*Corresponding author

Keywords: RNA-binding protein; *Xenopus*; zinc finger

Introduction

Protein-nucleic acid interactions are fundamental to the regulation of many biological processes. Most DNA-binding proteins recognize specific sequences in the context of the *B*-form helix, whereas most RNA-binding proteins utilize a combination of sequence and complex three-dimensional shapes to provide binding specificity (Steitz, 1993). An exception to the latter are the double stranded RNA-binding proteins (dsRBPs) which, like DNA-binding proteins, can interact with a single shape, the right handed helix. However, unlike DNA, an RNA helix is always *A*-form, with a very deep and narrow major groove and a wide and shallow minor groove (Delarue & Moras, 1989). As a consequence of the *A*-form geometry, functional groups in the major groove that can be used to distinguish different base-pairs are inaccessible to dsRBPs unless the helix is distorted. Hence, although many DNA-binding proteins have been demonstrated to bind specific sequences within the context of a right-handed helix, the dsRBPs studied so far show no signs of sequence specificity (Bass *et al.*, 1994; St Johnston *et al.*, 1992). Given the geometry of the *A*-form helix, dsRBPs are more likely to bind in the minor groove (Bevilacqua & Cech, 1996) where dsRNA molecules are chemically very similar and distinguishing one sequence from another is more difficult (Seeman *et al.*, 1976).

For many dsRBPs, the question of what substrates they recognize *in vivo* is unanswered. Although there are some naturally occurring antisense RNAs, and the presence of dsRNA in mammalian cells in the absence of viral infection has been reported (reviewed by Nicholson, 1996), the source and abundance of dsRNA in cells is essentially unknown. However, studies of dsRBPs for which substrates are known show that dsRBPs do not always require a completely base-paired substrate. For example, the natural substrates of RNase III (Court, 1993) and dsRNA adenosine deaminases (Melcher *et al.*, 1996; Polson *et al.*, 1996), although largely double stranded, are frequently interrupted by mis-matches, bulges and loops.

Recently, a dsRNA-binding motif (dsRBM) present in more than 18 different proteins was identified (Bass *et al.*, 1994; Gibson & Thompson, 1994; St Johnston *et al.*, 1992). Proteins containing dsRBMs bind dsRNA with high affinity and some

Abbreviations used: dsRBP, double stranded RNA-binding protein; ZFP, zinc finger protein; uORF, upstream open reading frame; UTR, untranslated region; PCR, polymerase chain reaction; CV, column volume; IPTG, isopropyl-β,D-thiogalactopyranoside; RACE, rapid amplification of cDNA ends; CV, column volumes.

0022–2836/97/320195–14 $25.00/0/mb971177

© 1997 Academic Press Limited

also bind RNA-DNA hybrids, but with lower affinity (Bass *et al.*, 1994). To identify new dsRBPs, we screened a *Xenopus* cDNA expression library with radiolabeled dsRNA.

Characterization of two dsRBM containing proteins identified in this screen was reported previously (Bass *et al.*, 1994). Here, we report the characterization of another cDNA identified in this screen that encodes a unique zinc finger protein. Our studies show this protein binds dsRNA and RNA-DNA hybrids with high affinity. Furthermore, consistent with studies of other dsRBPs, binding to either type of substrate is sequence independent.

The zinc finger, a repeated sequence motif first identified in the *Xenopus* transcription factor TFIIIA, is possibly the most abundant nucleic acid binding motif in nature. It has been estimated that up to 1% of the human genome encodes proteins containing zinc finger motifs (Hoovers *et al.*, 1992). Several types of zinc finger proteins (ZFPs) have been studied (Klug & Schwabe, 1995), including the C_2H_2 class as typified by TFIIIA. Although two C_2H_2 ZFPs, TFIIIA and p43, are involved in 5 S RNA storage (Joho *et al.*, 1990; Pieler & Bellefroid, 1994), most ZFPs bind DNA and act as transcription factors (Pieler & Bellefroid, 1994).

TFIIIA, unlike other C_2H_2 ZFPs, binds both DNA and RNA. However, deletion experiments reveal that the fingers binding DNA (fingers 1 to 3) are distinct from those binding RNA (fingers 4 to 6; Theunissen *et al.*, 1992). Besides TFIIIA, only eight C_2H_2 class ZFPs have been observed to bind a nucleic acid other than DNA and one of these is of synthetic design (Andreazzoli *et al.*, 1993; Joho *et al.*, 1990; Koster *et al.*, 1991; Shi & Berg, 1995). The three zinc fingers from the transcription factor Sp1, as well as the protein of synthetic design, can bind RNA-DNA hybrids (Shi & Berg, 1995). Nevertheless, naturally occurring non-DNA substrates have only been identified for the 5 S RNA binding proteins TFIIIA and p43. No previously identified ZFPs have been observed to bind dsRNA with high affinity, so we were surprised to find that our screening protocol identified a cDNA encoding a ZFP.

Results

Cloning of dsRBP-ZFa

In order to identify previously uncharacterized dsRBPs we screened a *Xenopus* ovary cDNA expression library in a manner analogous to that used for identifying DNA binding proteins (Singh *et al.*, 1988). Filters were probed with a radiolabeled dsRNA and four different clones were isolated (Bass *et al.*, 1994). Characterization of the binding properties and expression of two of these clones has been described (Bass *et al.*, 1994), and here we report the characterization of a 2062 nucleotide (nt) clone designated cDNA-24b.

A preliminary northern analysis using an *in vitro* synthesized transcript from cDNA-24b as a probe identified a single band, migrating between 2.6 and 2.8 kb, in *Xenopus* ovary poly(A)$^+$ RNA. Since cDNA-24b lacked an initiating methionine codon and was smaller than the band seen on the northern blot, it was used to screen another cDNA library to isolate a full-length clone. Using a *Xenopus* head cDNA library (Hemmati-Brivanlou *et al.*, 1991) we identified a 2514 nt clone designated cDNA-15a. Our criterion for a full-length clone was the presence of a stop codon in all three reading frames preceding the initiating methionine codon. cDNA-15a did not have a stop codon in the coding frame prior to a potential initiating codon, so we employed RACE (rapid amplification of cDNA ends) to obtain the 5′ end. Six of the longest RACE clones isolated were sequenced and of these, five ended at the same point and one was 10 nt shorter. We were convinced these represented the 5′ end of dsRBP-ZFa when we found that the sequence of each clone was identical and included a stop codon in the coding reading frame. cDNA-dsRBP-ZFa was constructed by ligating a 110 base-pair (bp) segment from one of the RACE clones onto the end of cDNA-15a. Both strands of cDNA-dsRBP-ZFa were sequenced.

cDNA-dsRBP-ZFa is 2625 nt in length and contains an open reading frame (ORF) of 1572 nt, a 216 nt 5′ untranslated region (UTR) and an 837 nt 3′ UTR. Within the ORF is a region of four 174 nt duplications (positions 877 to 1572) with a sequence identity ranging from 90 to 97%. The ORF begins with the second AUG found in the cDNA and the surrounding nucleotides (G**A**AU**AUG**-**G**CGGAC) fit the consensus sequence which includes a purine at position −3 and a G at position +4 (Kozak, 1987). Ninety two nt upstream of the second ORF is an AUG codon in a different reading frame and in a poor context (C**C**GG**AU**-**GU**CGCGA).

Protein sequence analyses

The predicted amino acid sequence of dsRBP-ZFa has several interesting features (Figure 1). The protein can be split into two main domains, a basic N-terminal domain (residues 1 to 432) with a predicted pI of 8.54, and an acidic C-terminal domain (residues 433 to 524, beginning with the asterisk in Figure 1) with a predicted pI of 3.98. Additionally, the C-terminal domain is rich in the three amino acids, serine (18%), proline (17%) and glycine (13%) (Creighton, 1984). A more detailed analysis of the charge distribution in the N-terminal domain reveals a periodic arrangement of acidic and basic regions bounded by proline residues.

The N-terminal domain is primarily composed of seven TFIIIA-like zinc finger motifs (underlined in Figure 1) and the linkers that join them. The zinc fingers generally follow the motif $Cx_2Cx_3\Phi x_5\Phi x_2Hx_5H$ where Φ is a hydrophobic residue, and x is any amino acid. The first three

```
MADEFGNGDALDLPVGKDAVNSLIRENSHIFS                                   32

                                   ooooooooooooooooooo
DTQCKVCSAVLISESQKLAHYQSRKHANKVRRYMAINQGEDSVPAKKFKAAPAEISDGEDR     93

SKCCPVCNMTFSSPVVAESHYIGKTHIKNLRLREQGGVKEGMVNQAKKTRTPTVATKSDNKMDHSDR   160

                          ooooo
AKFCKLCHSTFNNPLMAEQHYAGKKHKKQETKTQIMTIYTSSGQTPAQAPIPLNLNSPMPGSGSAGK   227

GFSCDKCNIVLNSIEQYQAHVSGAKHKNQLMSMTPLSEEGHQAVVAPSAIASGSAGK          284

GFSCDTCNIVLNSIEQYQAHISGAKHKNHLKSMTPLSEEGHTAAVAPSAFASGSAGK          341

GFSCDTCNIVLNSIEQYQAHISGAKHKNHLMSMTPLSEEGHTAAVAPSAFASGSAVK          398

                                   *
GFSCDTCNIVLNSIEQYQAHVSGAKHKNQLMSMTPSSEEGLPSAVGPSAFASPLSAG          455

GALSSGGPSGRGFCPSGDLTPKGPSSFGSLPPLGSLLPPLYPPAHSSQPYVHDDTMS          512

PDGYNYFNEDFE                                                       524
```

Figure 1. Predicted amino acid sequence of dsRBP-ZFa. The ORF shown begins with the second AUG codon of the cDNA. The sequence is arranged with the seven zinc finger motifs aligned to illustrate similarities between the finger motifs as well as between the linkers that join them. The last four fingers, underlined with filled black lines, are nearly identical to each other; the first three fingers, underlined with hatched lines, do not have strong similarities to each other or to the last four fingers. The beginning of the proline, serine and glycine rich acidic region is indicated with an asterisk. The putative nuclear localization signals (nls′) are shown with circles for the degenerate bipartite nls and squares for the overlapping SV40-like signals.

fingers are clearly distinct from the last four. In addition to being separated from the former by a short region containing five proline residues, the last four fingers, encoded by the DNA duplications mentioned above, are nearly identical repeats with 91 to 100% amino acid identity to each other.

The most striking difference between dsRBP-ZFa and other ZFPs is the length of the linker region that connects the fingers. In the vast majority of ZFPs this linker is fairly short, usually between six and eight amino acids (Berg, 1990; Hoovers *et al.*, 1992), whereas in dsRBP-ZFa it ranges from 34 to 44 amino acids.

Purification of recombinant dsRBP-ZFa

To obtain large quantities of protein for biochemical studies, c-DNA-dsRBP-ZFa was subcloned into the *Escherichia coli* T7 based expression vector pAED4 (Doering, 1992). Overexpressed dsRBP-ZFa was found exclusively in the insoluble fraction of lysed cells and attempts to increase its solubility were unsuccessful. Thus, we decided to solubilize the protein using guanidine HCl and refold it by dialysis as this method had been successful with other ZFPs (Fairall *et al.*, 1992). Electrophoresis of the refolded protein on a SDS-10% (w/v) polyacrylamide gel followed by silver staining revealed three major species (indicated by asterisks in Figure 2, S). The top band was full-length dsRBP-ZFa, the middle band was a truncation of dsRBP-ZFa missing residues from the C terminus, and the identity of the lower band was uncertain (data not shown; see Materials and Methods). We proceeded to purify full-length dsRBP-ZFa from the refolded protein using cation and anion exchange chromatography (see Materials and Methods). Purified dsRBP-ZFa (Figure 2, F) was stored as a 40% glycerol stock at −20° C. It should be noted that this preparation was essentially inactive in binding studies without addition of 5 to 10 mM DTT.

Mobility shift assays

Our initial library screen indicated that dsRBP-ZFa bound dsRNA but not DNA or ssRNA as the latter two were included in the hybridization mix as unlabeled competitors. It seemed likely that RNA-DNA hybrids would also be good candidate substrates for dsRBP-ZFa since they exhibit a con-

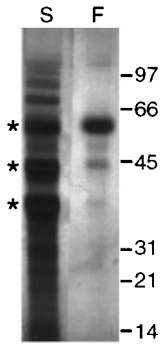

Figure 2. Purity of dsRBP-ZFa. Protein refolded from the insoluble fraction of *E. coli* cells was separated by SDS-10% PAGE and stained using silver. The starting material (S) and final purified preparation after ion exchange chromatography (F) are indicated. Asterisks indicate the three major species of staining material (see the text). In the final preparation, dsRBP-ZFa was determined by densitometry to be 78% of the total protein. Positions of molecular mass markers in kDa are shown to the right of the gel.

formation more similar to dsRNA than DNA (Gonzalez *et al.*, 1994; Salazar *et al.*, 1993). Furthermore, RNA-DNA hybrids were able to compete for dsRNA binding in studies of another dsRBP (Bass *et al.*, 1994).

We began our biochemical characterization of purified recombinant dsRBP-ZFa with gel mobility shift assays using five different substrates: a 36-mer dsRNA, two 36-mer RNA-DNA hybrids, a 100-mer dsRNA and a 100-mer RNA-DNA hybrid. The various 36-mer substrates were derived from the same sequence. The various 100-mer substrates were also derived from the same sequence (unrelated to the 36-mer sequence). Both 100-mer and 36-mer have no relationship to biologically relevant sequences.

Mobility shift assays were performed by mixing increasing amounts of purified dsRBP-ZFa with a ^{32}P-labeled substrate followed by electrophoresis on a native 6% polyacrylamide gel. Representative gels for each substrate are shown in Figure 3(a); to facilitate comparison, corresponding lanes on each gel have the same protein concentration. Mobility shift assays were done in triplicate for each substrate except for the 36-mer dsRNA which was repeated six times. The protein concentration required to shift about half of the substrate was similar for the two dsRNA substrates, while both hybrid substrates required slightly more protein. Using the data from the 36-mer dsRNA reactions, an apparent equilibrium dissociation constant (K_d) of 5.0×10^{-10} M was determined for the interaction corresponding to the first shift (Figure 3(b)). A similar analysis of the 36-mer RNA-DNA hybrid data yielded a K_d of 1.7×10^{-9} M. The latter was independent of which strand was DNA (data not shown). These K_d values assume that the protein used in the assays was fully active. To address this issue, we conducted titration experiments in which increasing amounts of dsRNA were mixed with a fixed concentration of dsRBP-ZFa. The results from preliminary experiments suggest the fraction of active protein was only 10 to 20% (data not shown), assuming the first shift represents a 1:1 stoichiometry of protein to RNA. Taking this into consideration, the K_d values listed above may be as much as an order of magnitude larger than the real equilibrium dissociation constants.

We observed a maximum of two shifts with 36-mer dsRNA and four with 100-mer dsRNA

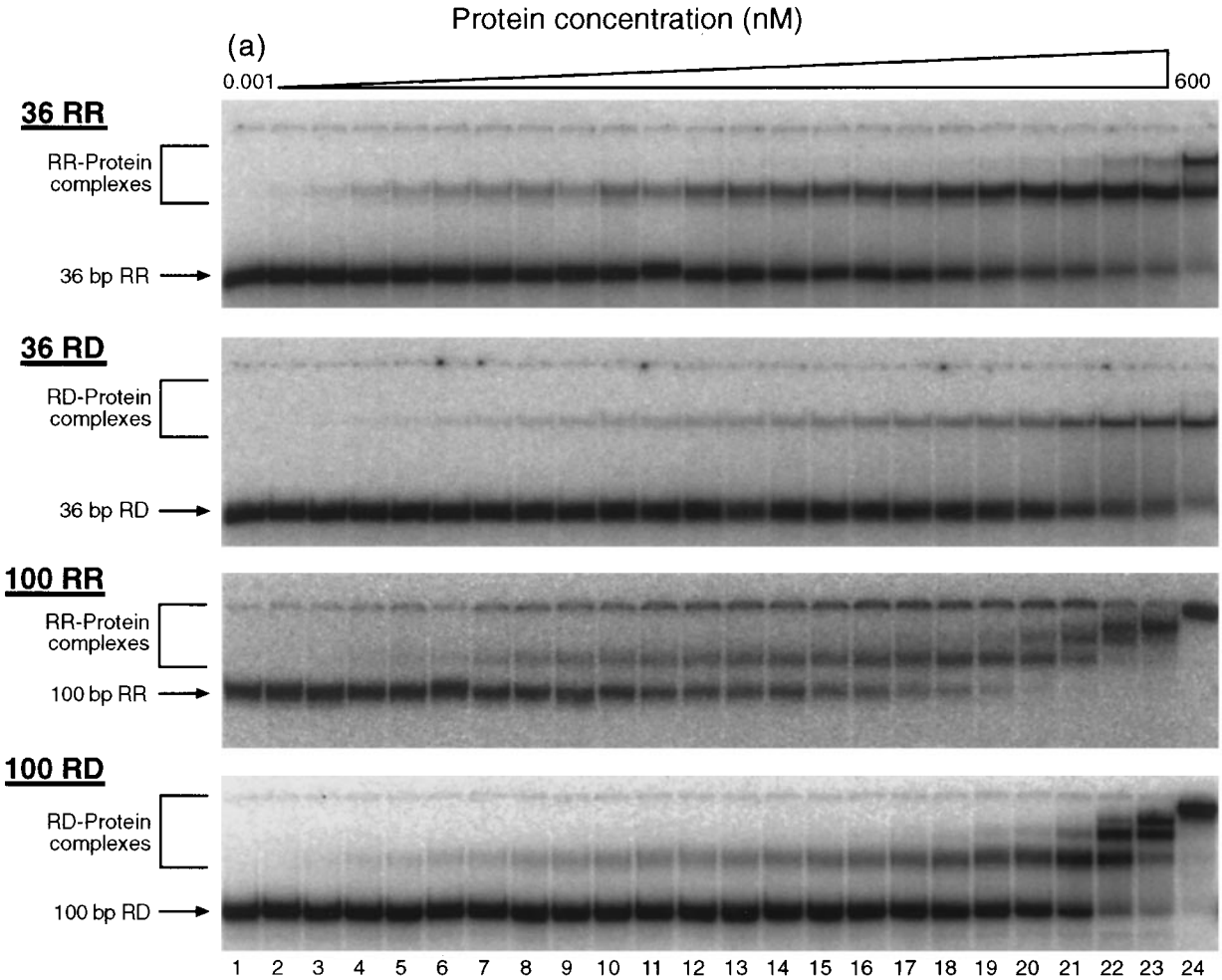

Figure 3(a) *(legend opposite)*

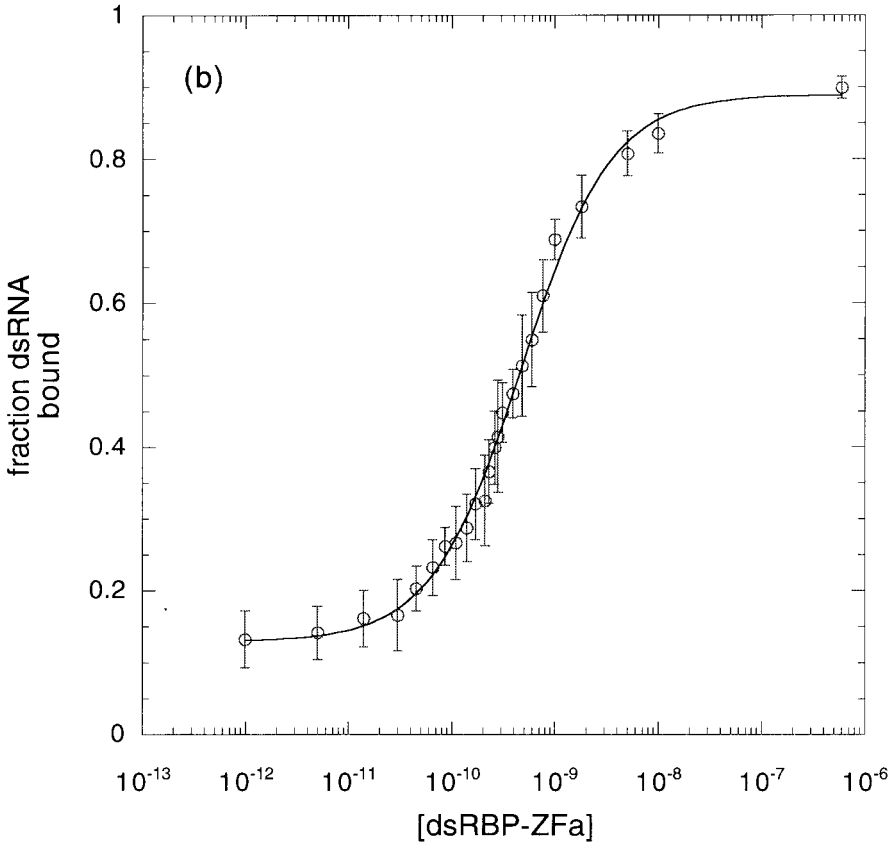

Figure 3. (a) Mobility shift analyses of dsRBP-ZFa binding to dsRNA and RNA-DNA hybrids. Either a 36-mer dsRNA (36 RR, 10 pM), a 36-mer RNA-DNA hybrid (36 RD, 10 pM), a 100-mer dsRNA (100 RR, 10 pM) or a 100-mer RNA-DNA hybrid (100 RD, 10 pM) was mixed with varying concentrations of dsRBP-ZFa. Corresponding lanes on each gel have the same concentration of dsRBP-ZFa to facilitate comparison of the affinities for the four different substrates. The protein concentration (M) in each lane is as follows: (1) 1.0×10^{-12}, (2) 5.0×10^{-12}, (3) 1.4×10^{-11}, (4) 3.0×10^{-11}, (5) 4.6×10^{-11}, (6) 6.6×10^{-11}, (7) 8.6×10^{-11}, (8) 1.1×10^{-10}, (9) 1.4×10^{-10}, (10) 1.7×10^{-10}, (11) 2.1×10^{-10}, (12) 2.3×10^{-10}, (13) 2.6×10^{-10}, (14) 2.8×10^{-10}, (15) 3.1×10^{-10}, (16) 3.9×10^{-10}, (17) 4.8×10^{-10}, (18) 6.0×10^{-10}, (19) 7.6×10^{-10}, (20) 1.0×10^{-9}, (21) 1.8×10^{-9}, (22) 5.0×10^{-9}, (23) 1.0×10^{-8}, (24) 6.0×10^{-7}. The electrophoretic mobilities of free and bound substrates are indicated in the PhosphorImager images of the four mobility shift assays. (b) K_d determination. Primary data for dsRBP-ZFa binding to 36-mer dsRNA was generated in multiple experiments and quantified. Each point is an average of six experiments and error bars represent ± standard deviation from the mean.

(Figure 3(a)). Ignoring the possibility of protein dangling, overlap of binding sites and the ability of only a portion of dsRBP-ZFa to bind dsRNA tightly, these data suggest a binding site size of 18 to 25 bp. We assume that initial binding is random along the entire length of the dsRNA and that the binding affinity for each site is identical. Although future studies will be necessary to verify this assumption, if correct, it is possible that binding can be described using conditional probabilities as proposed for the binding of non-sequence specific DNA binding proteins (McGhee & von Hippel, 1974) and other dsRBPs (Bass *et al.*, 1994).

Substrate preference

To further analyze the substrate preference of dsRBP-ZFa, tritiated or non-radioactive 100-mer RNA, RNA-DNA and DNA substrates were used as competitors of dsRBP-ZFa binding to ^{32}P-labeled 36-mer dsRNA. In seven individual experiments we analyzed the ability of each substrate to compete at 1× and 10× molar excess over 36-mer dsRNA by gel mobility shift assays (Figure 4(a)). These data, graphed in Figure 4(b), show that dsRNA competes best, in agreement with the higher K_d determined for RNA-DNA hybrids as compared with dsRNA. They also demonstrate that neither ssRNA nor DNA significantly compete for dsRNA binding, although ssRNA competes slightly better than DNA likely due to its limited ability to form intra and intermolecular duplexes.

Western analyses of dsRBP-ZFa expression

As a first step towards understanding the biological role of dsRBP-ZFa, polyclonal antibodies were raised in rabbits. cDNA-dsRBP-ZFa was subcloned into the *Escherichia coli* T7 based pET-16b expression vector (Novagen) and the resulting His-

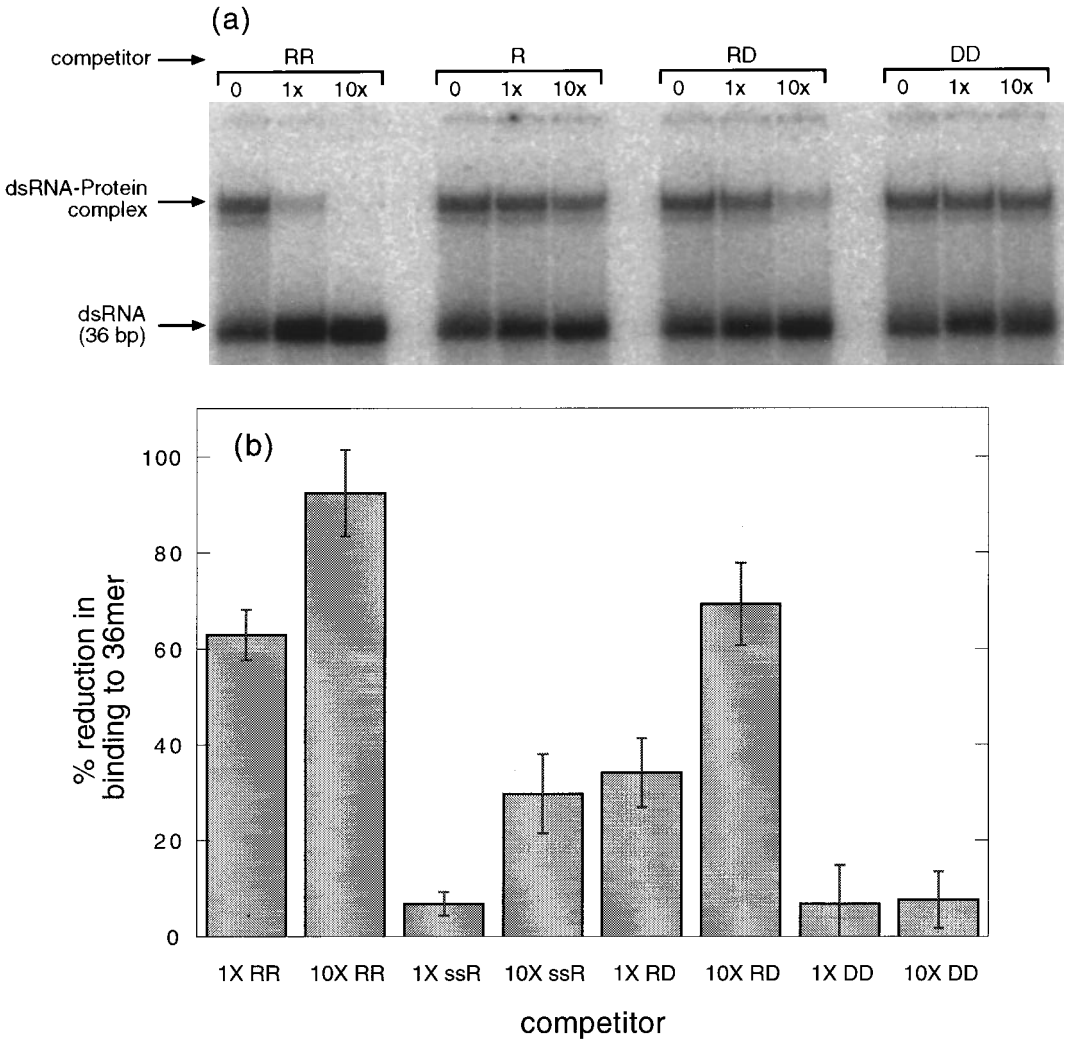

Figure 4. (a) Competition studies. A PhosphorImager image of a mobility shift competition assay is shown. The ability of four substrates (RR = 100 bp dsRNA, R = 100 bp ssRNA, RD = 117 bp RNA-DNA hybrid, DD = 143 bp dsDNA) to compete for dsRBP-ZFa binding to radiolabeled 36-mer dsRNA (10 pM) was assayed by mobility shift assays. dsRBP-ZFa was present at 0.5 nM in all lanes and competitors were added at 1× and 10× the concentration of 36-mer dsRNA. (b) The ability of each substrate to compete was determined as in (a) in four experiments and the average shown graphically here. Error bars represent ± standard deviation from the mean.

tagged protein purified by affinity chromatography for injection into rabbits. The protein produced from this vector was purposefully truncated to eliminate the C-terminal 89 amino acids and designated dsRBP-ZFaΔpro. We had initially produced polyclonal antibodies using the ORF from cDNA-24b, which included the C terminus, but these cross-reacted with oocyte keratin proteins which have a similar mobility as dsRBP-ZFa (data not shown).

The antibody produced with dsRBP-ZFaΔpro recognized a single polypeptide that was primarily localized to the nucleus (Figure 5A, N), although an immunoreactive polypeptide from cytoplasm became visible upon longer exposure (Figure 5B, C). The mobility of the immunoreactive protein was consistent with that of dsRBP-ZFa, but faster in whole oocytes and cytoplasm than in nuclei. We believe this was a result of the much larger quantity of protein loaded on the oocyte and cytoplasm lanes in order to maintain oocyte equivalents (see Materials and Methods).

Consistent with localization of dsRBP-ZFa to the nucleus, we identified two overlapping candidate nuclear localization signals starting at positions 184 (KKHK) and 185 (KHKK) that fit the four residue pattern for nuclear targeting (indicated by the squares in Figure 1; Dingwall & Laskey, 1991). We also identified a bipartite nls (Dingwall & Laskey, 1991) beginning at position 63 with a spacer length of 13 residues (**RR**YMAINQGEDSVPA**KKFK**, indicated by the circles in Figure 1). All three of the putative localization signals reside in the presumed binding domain of dsRBP-ZFa, something frequently observed in nucleic acid binding proteins (LaCasse & Lefebvre, 1995).

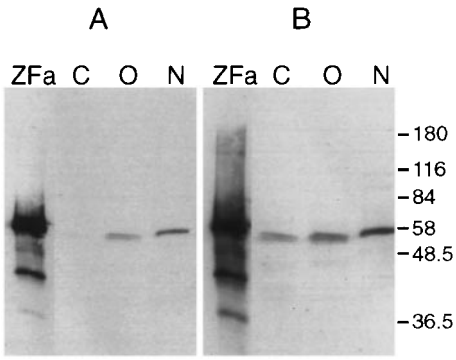

Figure 5. Western analysis. A polyclonal antibody against the zinc finger region of dsRBP-ZFa was used in a western analysis of dsRBP-ZFa expression in stage 6 *Xenopus* oocytes. Oocytes (O) were dissected into nuclear (N) and cytoplasmic (C) parts and proteins extracted and separated by SDS-10% PAGE. Purified dsRBP-ZFa (600 fmol; ZFa) was used as a marker. (B) is a longer exposure of (A).

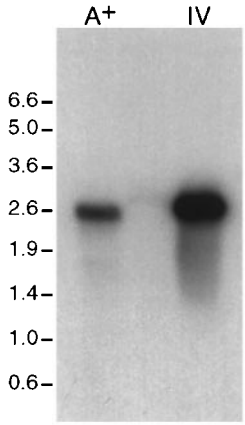

Figure 6. Northern analysis. Expression of dsRBP-ZFa mRNA was analyzed by probing a northern blot with a radiolabeled random primed probe generated from a 617 bp restriction fragment of the dsRBP-ZFa cDNA. A^+, 5 μg of poly(A)$^+$ RNA. IV, a 1700 nt *in vitro* hybridization control containing 0.5 ng of a synthetic transcript corresponding to the sense strand of dsRBP-ZFa cDNA (includes 36 nt of vector sequence). An autoradiograph of the northern blot is shown and mobilities of RNA markers (kb) are indicated on the left.

Northern analyses

We characterized dsRBP-ZFa mRNA expression in *Xenopus* ovary poly(A)$^+$ RNA by northern analyses. A ^{32}P-labeled random primed probe derived from cDNA-dsRBP-ZFa hybridized to a single band of approximately 2.6 kb (Figure 6, A^+). This closely matched the length of a 2762 nt transcript produced *in vitro* from cDNA-dsRBP-ZFa and electrophoresed on the same gel (Figure 6, IV). We estimate the amount of dsRBP-ZFa mRNA in an oocyte to be less than 2 attomoles by PhosphorImager quantitation, using a value of 80 ng poly(A)$^+$ RNA per *Xenopus* oocyte (Gurdon & Wickens, 1983). At approximately 1.2×10^{-6} copies per oocyte, the amount of dsRBP-ZFa is similar to TFIIIA (Ginsberg *et al.*, 1984) and several other mRNAs (Onate *et al.*, 1992; Reddy *et al.*, 1991; Taylor *et al.*, 1986), and thus appears to be of average abundance.

Discussion

We have identified a cDNA encoding a novel ZFP, dsRBP-ZFa, that binds dsRNA and RNA-DNA hybrids with high affinity. Along with seven zinc finger motifs, the protein has an acidic C terminus rich in proline, serine and glycine. Using gel mobility shift assays, we have characterized the nucleic acid binding properties of overexpressed, full-length protein. dsRBP-ZFa shows no apparent sequence specificity, but greatly prefers to bind *A*-form helices over other nucleic acids. dsRBP-ZFa binds dsRNA with an apparent K_d of 0.5 nM and RNA-DNA hybrids with an apparent K_d of 1.7 nM. dsRBP-ZFa is the first ZFP discovered that shows a preference for binding dsRNA. Perhaps the most significant aspect of the discovery of dsRBP-ZFa is that it shows there is another way to bind dsRNA other than using the previously characterized dsRNA binding motif (Bass *et al.*, 1994; Gibson & Thompson, 1994; St Johnston *et al.*, 1992) or sequences within other dsRBPs for which a common motif has not been defined (Denzler & Jacobs, 1994; Qian *et al.*, 1995; Qiu *et al.*, 1995).

Interesting features of the cDNA

A common feature of many ZFPs are apparent gene duplication events that result in multiple, contiguous zinc finger motifs in a single protein (Berg, 1990; Pieler & Bellefroid, 1994). The last four fingers of dsRBP-ZFa, encoded by four almost perfect 174 nt duplications, appear to be a result of such a phenomenon. In fact, obtaining the correct nucleotide sequence for this region of the cDNA would have been very difficult without automated sequencing technology as there are no unique priming sites over a 700 bp region.

Besides the ORF, the cDNA reveals other interesting features. Preceding the initiating AUG is a 162 nt overlapping upstream ORF (uORF). uORFs are found in many translationally regulated proteins (Geballe & Morris, 1994), although overlapping uORFs are rare in eukaryotic mRNAs. *In vitro* experiments with synthetic templates containing overlapping uORFs demonstrate that they significantly reduce translational efficiency (Peabody & Berg, 1986). Since the upstream AUG of dsRBP-ZFa is in poor context, lacking purine nucleotides at positions −3 and +4 (Kozak, 1987), leaky scanning could allow translation to initiate at the second AUG (Kozak, 1994). Database searches using the hypothetical polypeptide produced from the uORF did not identify any significantly similar

sequences. In addition to uORFs, long leader sequences can also affect protein expression through their ability to form secondary structures refractory to translation. dsRBP-ZFa has a 216 nt 5′ UTR, whereas a survey of 346 cDNAs found that the majority had 5′ UTRs that are shorter than 100 nt and only 8% were longer than 200 nt (Kozak, 1987).

The cDNA also appears to lack a good polyadenylation signal near the 3′ end. In fact, the only reasonable signal is the AAGAAA sequence located near the end of the cDNA (position 2516 in the cDNA), which is a very poor polyadenylation signal (Wickens, 1990). However, we were able to enrich for dsRBP-ZFa mRNA by poly(A)$^+$ selection, and all three of the cDNA clones we isolated ended in stretches of up to 33 adenosines. Perhaps future experiments will determine the ability of the AAGAAA sequence to stimulate polyadenylation of dsRBP-ZFa RNA.

dsRBP-ZFa has binding properties that are different from other ZFPs

Crystal structures have been solved for two ZFP transcription factors, Zif268 and Tramtrack, in complex with their DNA substrates (Fairall *et al.*, 1993; Pavletich & Pabo, 1991). The structures indicate that the proteins make contacts primarily with one strand of the DNA, and that the DNA is underwound relative to *B*-form DNA. Since the DNA strand in an RNA-DNA hybrid forms a structure intermediate between *A* and *B*-form (Gonzalez *et al.*, 1994; Salazar *et al.*, 1993), possibly similar to the observed underwound DNA, the ability of ZFPs to bind RNA-DNA hybrids has been studied. It has been observed that Sp1 and ZF-QQR, proteins that are similar to Zif268 and Tramtrack, can bind RNA-DNA hybrids, but the binding shows a strand preference. When the strand analogous to that contacted the most in the crystal structures is DNA, tighter binding is observed (Shi & Berg, 1995). Other studies demonstrate that not all ZFPs can bind RNA-DNA hybrids. For example, binding studies using TFIIIA and RNA-DNA hybrids constructed using 5 S ICR sequences reveal only nonspecific, low affinity interactions (Nightingale & Wolffe, 1995).

Binding of dsRBP-ZFa to RNA-DNA hybrids differs from that of Sp1 and ZF-QQR. First, dsRBP-ZFa does not show any sequence specificity among the substrates tested, whereas Sp1 and ZF-QQR only bind to specific sequences (Shi & Berg, 1995). Second, dsRBP-ZFa had the same affinity for the two 36-mer RNA-DNA hybrids, independent of which RNA strand was used to prepare the RNA-DNA substrate. The latter is particularly notable since the two strands of RNA that comprise the 36-mer dsRNA differ such that one strand is rich in purine bases and the other in pyrimidine bases. Finally, whereas dsRBP-ZFa binds dsRNA with a higher affinity than it binds RNA-DNA hybrids, Sp1 and ZF-QQR have negligible affinity for dsRNA (Shi & Berg, 1995).

dsRNA binding by dsRBP-ZFa is also quite different from 5 S RNA binding by TFIIIA and p43. The interaction of the latter two proteins with 5 S RNA primarily depends upon the complex shape of 5 S RNA (reviewed by Pieler & Bellefroid, 1994), whereas dsRBP-ZFa prefers the relatively simple *A*-form helix.

In sum, there appears to be a fundamental difference between how dsRBP-ZFa and other ZFPs recognize their substrates. Taken together, our results suggest that the substrate conformation, rather than its sequence, is paramount for substrate recognition by dsRBP-ZFa. We have recently determined that dsRBP-ZFaΔpro, a truncated protein which lacks the C-terminal acidic region, binds dsRNA at least as tightly as the full-length protein (data not shown; see Materials and Methods). Thus, as might be expected, the unique binding properties of dsRBP-ZFa appear to derive from the N-terminal zinc finger domain, rather than the acidic C-terminal domain. We have also observed that binding of dsRBP-ZFa to dsRNA is inhibited by EDTA, consistent with the involvement of the zinc fingers in nucleic acid binding (data not shown).

dsRBP-ZFa has unusual zinc finger motifs

In the classical zinc finger motif, coordination of zinc by two cysteine and two histidine residues establishes the structure of the finger. This structure is further stabilized by a hydrophobic core comprised of a conserved aromatic residue at position +4 and a conserved aliphatic residue at position +10 (Figure 7). NMR studies of classical fingers reveal that the aromatic residue at position +4 stacks on the first zinc coordinating histidine to form one part of the hydrophobic core (Jasanoff & Weiss, 1993). None of the fingers in dsRBP-ZFa are perfect matches to the consensus for the classical zinc finger motif (Figure 7; Jacobs, 1992). While the first finger lacks a hydrophobic residue at position +10 and is the most dissimilar to the consensus overall, the second and third fingers, with suitable hydrophobic residues at positions +4 and +10, are most similar. The last four fingers of dsRBP-ZFa have hydrophobic residues at the correct positions, but the location of the aromatic and aliphatic residues at positions +4 and +10 are reversed from the consensus, and tyrosine is found instead of phenylalanine (Figure 7).

Prior to our characterization of dsRBP-ZFa, we observed dsRNA binding using a truncated protein produced from the ORF of cDNA-24b (ORF beginning at isoleucine 115 in Figure 1; data not shown; see Materials and Methods). Since 24b lacks the first two zinc fingers of dsRBP-ZFa, this observation showed that these amino-terminal fingers were not required for binding. It is enticing to speculate that the unusual arrangement of amino acids in the last four fingers dictates the preference for binding to *A*-form helices. Of course, it is also possible that the unusually long linkers are directly

```
                          +4    +10
                           |     |
Classical        F  C  C   F     L   H     H
dsRBP-ZFa:1  FSDTQ CKVC SAVLISESQKLA HYQSRKH ANKVR
dsRBP-ZFa:2  DRSKC CPVC NMTFSSPVVAES HYIGKTH IKNLR
dsRBP-ZFa:3  DRAKF CKLC HSTFNNPLMAEQ HYAGKKH KKQET
dsRBP-ZFa:4  GKGFS CDKC NIVLNSIEQYQA HVSGAKH KNQLM
dsRBP-ZFa:5  GKGFS CDTC NIVLNSIEQYQA HISGAKH KNHLK
dsRBP-ZFa:6  GKGFS CDTC NIVLNSIEQYQA HISGAKH KNHLM
dsRBP-ZFa:7  VKGFS CDTC NIVLNSIEQYQA HVSGAKH KNQLM
```

Figure 7. Zinc finger alignment. The zinc finger motifs from dsRBP-ZFa are shown. The consensus for the classical motif, determined using 1340 fingers from 221 ZFPs (Jacobs, 1992), is printed above the motifs from dsRBP-ZFa. Numbers given are relative to the second cysteine residue involved in zinc coordination and spaces in the sequences are present merely to facilitate comparison.

responsible for binding nucleic acid, and the zinc fingers function to position the linkers in a conformation required for binding.

Among known ZFPs, dsRBP-ZFa is most similar to a group that contain widely spaced finger motifs. Unlike the majority of ZFPs that have motifs separated by a six to eight residue linker whose sequence is highly conserved, these ZFPs have motifs separated by 7 to 410 residues that do not define a consensus. Previously, four proteins, *tsh*, TRS-1, *suvar(3)7*, and *Pep*, were reported to be members of this family (Fasano *et al*., 1991). Interestingly, also like dsRBP-ZFa, many of the zinc fingers in these proteins have an interhistidine distance of five residues; possibly this interhistidine distance is somehow correlated with long linkers. Based upon these similarities, dsRBP-ZFa and human alpha-fetoprotein enhancer binding protein (GenBank D10250; Morinaga *et al*., 1991), a protein frequently seen in database searches using dsRBP-ZFa, appear to be members of this group, bringing the total to six examples of ZFPs in this family.

Possible biological roles of dsRBP-ZFa

Western analyses of dsRBP-ZFa expression showed that the protein is primarily in the nuclei of *Xenopus* oocytes, with lower, but detectable, levels in the cytoplasm (Figure 5). Thus, at present, biological functions requiring a nuclear localization, as well as those requiring a cytoplasmic localization, can be considered. Database searches using the predicted amino acid sequence of dsRBP-ZFa identified four similar sequences: three human EST clones (GenBank F06769, TIGR EST19286 and THC116584) and a clone derived from a human X chromosome (GenBank HSJ3E10). The three EST clones appear unrelated to each other and show similarities with parts of the first three zinc fingers in dsRBP-ZFa, whereas clone HSJ3E10 is similar to the last finger in dsRBP-ZFa. Unfortunately, none of these sequences have a known biological function so they do not help elucidate the cellular role of dsRBP-ZFa.

Overall, the primary structure of dsRBP-ZFa is reminiscent of a transcription factor: it contains a basic N-terminal domain with seven zinc finger motifs and an acidic C terminus rich in proline, serine and glycine. Some DNA sequences are known to assume an *A*-form conformation in solution (Basham *et al*., 1995). Since dsRBP-ZFa prefers to bind *A*-form nucleic acids, possibly a DNA sequence with an *A*-form conformation could be a substrate for dsRBP-ZFa, and dsRBP-ZFa could act as a transcription factor.

The C terminus of dsRBP-ZFa, rich in proline, serine and glycine residues, has weak similarities to proteins that are also rich in these amino acids. Proline rich regions are often involved in interactions with other proteins; possibly this region of dsRBP-ZFa mediates similar interactions. For instance, in some proteins proline rich regions act as transcriptional activation domains (Mitchell & Tjian, 1989) that interact with RNA polymerase or an associated factor to target the polymerase to specific genes (Ptashne, 1988). Additionally, Src homology 3 (SH3) domains bind proline rich regions of proteins that function in signal transduction (Yu *et al*., 1994). Although the C terminus of dsRBP-ZFa contains two regions (PKGP and PLYP) that fit the loosely defined motif for SH3 recognition (PXXP), neither of these fit a more defined motif ($X_P\Phi P_PXP$, where X is any amino acid, Φ is a hydrophobic amino acid and lower case letters indicate a weaker preference at that position; Yu *et al*., 1994).

In addition to the possible roles suggested by its protein sequence, many other possible biological functions can be imagined. Previously characterized dsRBPs have a wide variety of functions, including RNA localization (St Johnston *et al*., 1992), participation in the antiviral response (Nicholson, 1996), RNA processing (Court, 1993) and RNA editing (Melcher *et al*., 1996; Polson *et al*., 1996). Although it is hard to imagine that dsRBP-ZFa has catalytic activity like the dsRBPs involved in RNA processing and editing, functions analogous to non-catalytic dsRBPs are certainly possible. Additionally, dsRBP-ZFa could play a regulatory role by binding base-paired structures formed during RNA splicing, or to the RNA-DNA hybrids known to occur during transcription and DNA replication. Conceivably, dsRBP-ZFa could also act to protect dsRNA from modification by dsRNA adenosine deaminase, as observed for a factor in the cytoplasm of *Xenopus* oocytes (Saccomanno & Bass, 1994). Hopefully, future studies will lead to an understanding of the biological function of dsRBP-ZFa, and thus expand our general knowledge of the biological roles of dsRBPs. In addition, the discovery of dsRBP-ZFa presents a unique opportunity to understand, in future studies, how the zinc finger motif can be adapted to bind substrates other than DNA.

Materials and Methods

Oligonucleotides and plasmids

Oligonucleotides used as probes were gel purified on 20% acrylamide, 8 M urea gels then labeled with phage T4 polynucleotide kinase and [γ-^{32}P]ATP. Oligonucleotides used as primers in PCR were generally not gel purified unless they also happened to be used as hybridization probes. The oligonucleotides used were: (R8) 5′-GGGGACGCGTTGGATTTG-3′; (F10) 5′-CCTCCCCCTGGTTTATCG-3′; (T3) 5′-AATTAACCCTCACTAAAGGG-3′; (SP6) 5′-GATTTAGGTGACACTATAG-3′; (RACE1) 5′-TGTTTTTTTTGCTTGGTTCA-3′; (F11) 5′-CAUCAUCAUCAUCTGGCGACTGCAGAGAGAGCG-3′; (R10) 5′-ATTCATTGTCCCGGATGTCG-3′; (zf-R11) 5′-CAGCACAAGCTCCTATTCC-3′; (zf-F14-del) 5′-CTGACGGATCCCCTTCCTAAGACGATGGTG-3′; (Met1) 5′-GGTTTTTAGGAAGCATATGGCGGACG-3′.

The following plasmids were purified using CsCl/ethidium bromide equilibrium centrifugation as described (Ausubel *et al.*, 1990). pGEM-13 zf + (Promega; used to subclone cDNA-24b); pBS SK- (obtained from *in vivo* excision of the *Xenopus* head library; Hemmati-Brivanlou *et al.*, 1991); pAMP1 (Life Technologies, Inc.; used to clone RACE products); pET-16b (Novagen; used to express dsRBP-ZFaΔpro); pAED4 (used to express dsRBP-ZFa; Doering, 1992).

Restriction fragments used for subcloning were gel purified on 1% (w/v) agarose/TAE gels, and isolated using the GLASSMAX DNA Isolation Spin Cartridge System (Life Technologies, Inc.).

Cloning and sequencing of cDNAs

cDNAs were sequenced either using a Sequenase 2.0 kit (USB; cDNA-24b, 15a and race1) or by the University of Utah Health Sciences Sequencing Facility.

The partial clone 15a was identified by screening a *Xenopus* cDNA library (Hemmati-Brivanlou *et al.*, 1991) as described (Bass *et al.*, 1994). Filters were incubated at 42°C for two to three hours in prehybridization solution (0.8× SSC (SSC is 0.15 M NaCl, 0.015 M $Na_3C_6H_5O_7$) 5× Denhardts, 0.05% (w/v) $Na_4P_2O_7$, 0.5% (w/v) SDS, 100 μg ml^{-1} salmon sperm DNA), then at 42°C for 14 to 16 hours with hybridization solution (0.8× SSC, 1× Denhardts, 0.05% $Na_4P_2O_7$, 100 μg ml^{-1} salmon sperm DNA) that contained a radiolabeled oligonucleotide (R8) then washed in 0.3× SSC, 0.05% $Na_4P_2O_7$ four times for 15 minutes at room temperature.

The phage clone containing the longest cDNA (15a), identified by polymerase chain reaction (PCR) using primers F10 and T3 (complementary to the vector), was excised from the phage using the Exassist^{TM} Interference-Resistant Helper Phage system (Stratagene) to yield pBS-15a (15a in pBS SK-).

RACE

The 5′ RACE System for Rapid Amplification of cDNA Ends kit (Life Technologies, Inc.) was used to obtain the 5′ end of dsRBP-ZFa. Procedures were performed according to instructions except that cDNA was prepared by reverse transcription of 200 ng of *Xenopus* oocyte poly(A)$^+$ RNA using a primer complementary to 15a (RACE1). The initial PCR used primer F10 and a 1:10 mixture of anchor primer and primer UAP provided with the kit. Secondary PCR was with primers F11 and UAP. Products from the PCR were cloned according to the kit and the resulting bacterial colonies screened as described (Sambrook *et al.*, 1989). Filters were incubated at 42°C for one hour in prehybridization solution (1× SSC, 5× Denhardts, 1% SDS, 0.1 mg ml^{-1} salmon sperm DNA), then at 42°C for six hours in hybridization solution (1× SSC, 1× Denhardts, 0.05% $Na_4P_2O_7$, 0.1 mg ml^{-1} torula yeast RNA) that contained a radiolabeled oligonucleotide (R10) then washed twice for 15 minutes at room temperature in 1× SSC. Positive colonies were further screened by PCR to determine the size of the inserts using a vector specific primer (SP6) in combination with a dsRBP-ZFa specific primer (F11); the six longest cDNAs were sequenced.

Construction of a full length clone of dsRBP-ZFa

pBS-15a DNA and pAMP-race1 were digested with *Not*I and *Nru*I (NEB). The 5436 bp (from pBS-15a) and 219 bp (from pAMP-race1) fragments were gel purified and ligated to yield pBS-ZFa. 110 bp of the dsRBP-ZFa sequence in this construct was derived from the pAMP-race1 clone and the rest from 15a.

Cloning, expression and purification of dsRBP-ZFaΔpro

pET-ZFaΔpro is the N-terminal 437 residues of dsRBP-ZFa in the T7 based expression vector pET-16b. As an intermediate to pET-ZFaΔpro, pAED-ZFaΔpro was first constructed. pAED-ZFa DNA (see next section) was digested with *Bst*EII and *Bam*HI (NEB). PCR on pAED-ZFa, using primers zf-R11 and zf-F14-del, was used to generate a stop codon at amino acid 436 and a *Bam*HI site 3′ of the new stop codon. The 708 bp PCR product was digested with *Bst*EII and *Bam*HI, the fragment gel purified and ligated to the 4160 bp pAED-ZFa *Bst*EII-*Bam*HI fragment. DH5α cells were transformed with the ligation product and DNA from a transformant containing the correct construct was used to build pET-ZFaΔpro.

pET-ZFaΔpro was constructed by digesting pAED-ZFaΔpro and pET-16b DNA with *Nde*I (Boehringer Mannheim) and *Bam*HI. The 1315 bp dsRBP-ZFa fragment from pAED-ZFaΔpro was ligated to the 5701 bp pET-16b fragment and the product (pET-ZFaΔpro) used to transform DH5α cells.

Protein was overexpressed as described in the pET manual using 1 l of media containing BL21(DE3)/pLys S cells transformed with pET-ZFaΔpro DNA. SDS-PAGE analyses of the soluble and insoluble fractions of lysed cells revealed an equivalent distribution of dsRBP-ZFaΔpro protein. To make use of all the protein, whole cells were solubilized in 80 ml of AF binding buffer (6 M urea, 5 mM imidazole, 32 mM Tris (pH 7.9), 0.5 M NaCl) using a vortex mixer and sonication. Insoluble material was removed by centrifugation at 40,000 *g* and the solution was further clarified by passage through a 0.2 μ filter before being subjected to Ni^{2+} affinity chromatography. Ni^{2+} affinity chromatography was carried out using an 8 ml column (HR 10/10, Pharmacia) packed with AF-Chelate-650-M resin (TosoHaas) charged with $NiSO_4$ and equilibrated in AF binding buffer attached to a Biocad Sprint Perfusion Chromatography System (Perseptive Biosystems). Solubilized material (40 ml) was loaded onto the column at 5 ml min^{-1} followed by 10 column volumes (CV) of 98.5% (v/v) AF binding buffer plus 1.5% (v/v) AF elution buffer (6 M

urea, 1 M imidazole, 20 mM Tris (pH 7.9), 0.5 M NaCl). Bound protein was eluted using a 0 to 100% gradient of AF elution buffer over 5 CV at 5 ml min^{-1} followed by 3 CV of 100% elution buffer into 1 ml fractions. Fractions containing dsRBP-ZFaΔpro were pooled, dialyzed against AF binding buffer, and re-chromatographed. Fractions containing dsRBP-ZFaΔpro were pooled again and the protein refolded by dialysis against 4 × 500 ml of AF refolding buffer (50 mM KPO_4 (pH 7.0), 100 μM $ZnCl_2$, 2.5 mM DTT). Protein concentration was determined by absorbance at 280 nm using a theoretical extinction coefficient of 1 $A_{280} = 3.54$ mg ml^{-1}.

Affinity-purified dsRBP-ZFaΔpro protein (3 mg) was gel purified for production of polyclonal antibodies in rabbits using SDS-10% polyacrylamide gels under reducing conditions. Protein bands were visualized for excision by soaking the gel in a solution containing 0.25 M KCl and 1 mM DTT which results in the precipitation of SDS surrounding the protein (Nelles & Bamburg, 1976). SDS was removed from the gel slices by washing three times in water and the slices sent to the Berkeley Antibody Company (BabCo) for production of rabbit polyclonal antibodies.

Cloning, expression and purification of dsRBP-ZFa

dsRBP-ZFa was subcloned into the T7 based expression vector pAED4. pBS-15a was digested with *Bam*HI and *Hin*dIII. The 1550 bp dsRBP-ZFa *Bam*HI-*Hin*dIII fragment was gel purified. pAED4 DNA was digested with *Bam*HI and *Nde*I and the 3189 bp fragment gel purified. Primers Met1 and F10 were used in PCR of pBS-15a to generate a 233 bp product which was digested with *Nde*I and *Hin*dIII to yield 146 and 87 bp fragments. The 146 bp PCR fragment was gel purified and a three-way ligation performed with the 1550 bp (dsRBP-ZFa) and 3189 bp (pAED4) fragments to produce pAED-ZFa.

Protein expression was as described for dsRBP-ZFaΔpro except that 3 l of media was inoculated with 3 ml of an overnight culture and isopropyl-β,D-thiogalactopyranoside (IPTG) was added to 0.4 mM. Cells were lysed with two rounds of freeze-thawing and sonication after each thaw, the insoluble fraction isolated by centrifugation at 40,000 *g* and the pellet washed three times by sonication in TGMK buffer (25 mM Tris (pH 7.9), 10% (v/v) glycerol, 12.5 mM $MgCl_2$, 100 mM KCl, 0.1% (v/v) NP-40). The pellet was dissolved in 200 ml of KPMKDZ buffer (50 mM $KHPO_4$ (pH 7.0), 5 mM KCl, 1 mM $MgCl_2$, 5 mM DTT, 100 μM $ZnCl_2$) and 8 M guanidine HCl using a vortex mixer and sonication. Material that would not dissolve was removed by centrifugation at 40,000 *g* and dissolved protein diluted to 250 ml with water, placed into dialysis tubing (Spectra/Por 3; Baxter) and dialyzed against 4 × 500 ml of KPMKDZ buffer for four hours at 4°C. Dialysis was followed by centrifugation at 40,000 *g* and the solution further clarified by passage through a 0.2 μ filter.

The clarified solution (260 ml) was adjusted to 150 mM NaCl and subjected to cation exchange chromatography using a 1.66 ml HS column (Perseptive Biosystems) equilibrated in HS binding buffer (KPMKDZ + 150 mM NaCl). Refolded protein (36 ml) was loaded onto the column at 5 ml min^{-1} followed by 40 CV of HS binding buffer. Protein was eluted using a 0.15 to 1 M NaCl gradient over 15 CV at 5 ml min^{-1} into 1 ml fractions. dsRBP-ZFa fractions were pooled, dialyzed against 4 × 4 l of EKMDZ buffer (50 mM ethanolamine (pH 9.0), 5 mM KCl, 1 mM $MgCl_2$, 5 mM DTT, 100 μM $ZnCl_2$) and subjected to anion exchange chromatography using a 1.66 ml HQ column (Perseptive Biosystems) equilibrated in EKMDZ buffer. HS eluate (15 ml) was loaded onto the column at 5 ml min^{-1} followed by 10 CV of EKMDZ buffer. Protein was eluted using a 0 to 1 M NaCl gradient over 15 CV at 5 ml min^{-1} into 1 ml fractions. Fractions containing dsRBP-ZFa were pooled, dialyzed against 8 × 500 ml of 20 mM Hepes-KOH (pH 7.0), 20 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 100 μM $ZnSO_4$ for four hours each at 4°C, and the resulting 60 ml concentrated to 6 ml using a stirred ultrafiltration cell (Amicon, model 8050) with a PM-30 Diaflo ultrafiltration membrane (Amicon). Concentrated dsRBP-ZFa was stored at −20°C after mixing with an equal volume of 80% glycerol, 20 mM Hepes-KOH (pH 7.0), 20 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 100 μM $ZnSO_4$.

The purity of dsRBP-ZFa protein was determined by densitometry of a silver stained gel and the relative amounts of each species determined using software provided (BioRAD Scanning Densitometer). The 55.6 kDa dsRBP-ZFa band was determined to be 78% of the total protein.

Prior to storage, dsRBP-ZFa protein concentration was determined by measuring absorbance at 280 nm using a theoretical extinction coefficient of 1 $A_{280} = 3.0$ mg ml^{-1} and at 205 nm using a theoretical extinction coefficient of 1 $A_{205} = 0.035$ mg ml^{-1} (Scopes, 1987). Both wavelengths gave similar results and values were averaged. The final concentration used to make dilutions for gel shift reactions was obtained by multiplying the concentration obtained by spectrophotometry (2.25 mg/ml, 17.3 μM) by the purity determined by densitometry (78%) to yield a value of 1.76 mg ml^{-1} or 13.5 μM dsRBP-ZFa protein in the 40% glycerol stock.

All three of the major species present prior to purification of dsRBP-ZFa (Figure 2, lane S) reacted with anti-dsRBP-ZFa antibodies, but only the top and middle bands bound a Ni^{2+} affinity column when dsRBP-ZFa was expressed with an N-terminal histidine tag; the latter suggests the middle band was a C-terminal truncation.

Mobility shift assays

dsRNA (36-mer) was prepared as described (Polson & Bass, 1994) except single strands were first annealed then treated with calf intestinal phosphatase and labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP.

RNA-DNA (36-mer) hybrids (36 bp) were prepared by treating 36-mer ssRNAs with calf intestinal phosphatase, mixing equimolar amounts of the ssRNA and the appropriate 36-mer DNA oligo in TE, heating to 65°C for five minutes, slow cooling to 30°C over an hour then incubating on ice for five minutes. Hybrids were gel purified by 20% native PAGE and quantified, then labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP.

dsRNA (100-mer) was prepared as described (Bass *et al.*, 1994). 100-mer RNA-DNA hybrid (117 bp) was prepared by mixing 300 pmol of a T3 primer with 100 pmol of the 100-mer T7 RNA polymerase transcript in 50 mM KCl and 25 mM Hepes (pH 7.8), heating to 100°C for five minutes, slow cooling to 30°C over an hour then incubating on ice for five minutes. After ethanol precipitation the primer was extended using SuperScript RT II (Life Technologies, Inc.) at 42°C for one hour. After one hour an additional aliquot of reverse transcriptase was added and the incubation continued for another hour.

The hybrid was purified by 8% native PAGE, quantified and labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. All putative hybrids were verified as base-paired RNA-DNA strands by sensitivity to RNase H.

Mobility shift assays were performed by mixing all components of the reactions in 1 × GSB (20 mM Hepes-KOH (pH 7.0), 20 mM KCl, 1mM $MgSO_4$, 10 μM $ZnSO_4$, 10 mM DTT, 10% glycerol, 100 μg/ml BSA, and 0.1% NP-40). Protein was diluted in 1 × GSB when necessary. Binding reactions incubated over a 5 to 30 minutes time course at 25°C demonstrated no change in binding, so for convenience binding reactions were incubated at 25 to 28°C for 15 minutes. Binding reactions were stopped by loading onto a native 6% polyacrylamide gel. No dyes were included in the binding reactions as preliminary experiments indicated they inhibited binding.

Gels were 18 cm in length and composed of 6% acrylamide prepared using a 40% acrylamide stock containing an acrylamide:bis-acrylamide ratio of 19:1 (BioRAD), 0.5 × TB buffer (10× is 1 M Tris Base, 0.83 M H_3BO_3), 5% glycerol and 5 mM DTT and run in 0.5 × TB buffer. Gels were pre-run 1 to 2 hours at 4°C at 150 volts. Samples were subjected to electrophoresis for 2.5 hours at 150 volts, the gels dried and subjected to PhosphorImager analyses. Quantitation of the mobility shift assays and binding curve analyses was performed as described (Bass *et al.*, 1994) except that KaleidaGraph (Abelbeck Software) was used.

The fraction of active protein was determined in mobility shift assays in which 5.0×10^{-8} M dsRBP-ZFa was titrated with increasing amounts of dsRNA. The amount of active protein was determined by multiplying the concentration of RNA added to the reaction by the fraction shifted.

The dsRBP-ZFaΔpro protein used in mobility shift assays differed from the full length protein in that it was a fresh preparation that had not been stored at −20°C or mixed with glycerol. We did not determine the fraction of active dsRBP-ZFaΔpro protein in this preparation. Reaction conditions for mobility shift assays with dsRBP-ZFaΔpro were the same as for the full length protein except that the concentration of dsRNA was lowered to 1×10^{-12} M to account for the observed lower K_d ($\sim 2 \times 10^{-11}$ M for 36-mer dsRNA). Due to the differences in protein preparation and the length of storage for the full length protein *versus* dsRBP-ZFaΔpro it was not possible to make a direct comparison of the K_d values obtained for the two proteins.

In preliminary mobility shift assays using protein produced from the ORF in cDNA-24b, we determined a K_d of $\sim 3 \times 10^{-9}$ M for 36-mer dsRNA. However, both the reaction conditions and gel composition differed from those used with dsRBP-ZFa and the fraction of active 24b protein was not determined.

Immunoblotting

Oocytes were dissected into nuclear and cytoplasmic parts and proteins extracted as described except that dissections were performed in MBS not TGKED buffer (Saccomanno & Bass, 1994). Proteins were separated by SDS-10% PAGE and transferred to nitrocellulose using a semi-dry electroblotting apparatus. Western blots were performed using enhanced chemiluminescence (Amersham). Anti-dsRBP-ZFaΔpro serum was diluted 1:500, bound antibody was detected using an anti-rabbit IgG/peroxidase conjugate diluted 1:10,000 and washing was with 0.2% NP-40 in phosphate-buffered saline (PBS).

Northern analyses

Xenopus total and poly(A)$^+$ RNA were prepared as described (proteinase K method; Ausubel *et al.*, 1990). Electrophoresis was as described (Tsang *et al.*, 1993) and RNA transferred to a GeneScreen (NEN) membrane using a semi-dry electroblotting apparatus at 100 mA for one hour. A random primed probe was generated using a Random Primed DNA Labeling kit (Boehringer Mannheim) with a gel purified 617 bp restriction fragment prepared by digesting pBS-ZFa DNA with *Bst*EII and *Bgl*II (NEB). The probe was passed over a Chroma spin-30 column to remove unincorporated counts. Prehybridization and hybridization were as described (Ausubel *et al.*, 1990). After hybridization the membrane was washed at room temperature as follows: 6 × SSC for five minutes, 2 × SSC + 0.1% SDS for 25 minutes, 1 × SSC + 0.1% SDS for 15 minutes, 0.5 × SSC + 0.1% SDS for ten minutes, 0.2 × SSC + 0.1% SDS for ten minutes. Quantitation of the blot was performed using a PhosphorImager.

Sequence analyses

Protein sequence analyses were conducted using the LaserGene package from DNASTAR and the GCG package. The four residue pattern for nuclear localization was identified using the PSORT server (http://psort.nibb.ac.jp/; Nakai & Kanehisa, 1992). The degenerate bipartite nls was identified using the bipartite nls locator (http://cymac.welc.cam.ac.uk/biploc_form.html). Searches of public databases were conducted using the Linux versions of blastp and tblastn (Altschul *et al.*, 1990) provided by National Center for Biotechnology Information. The Institute for Genomic Research database was also used to identify similar clones.

Accession number

The GenBank data bank accession number for *Xenopus* dsRBP-ZFa cDNA is AF005083.

Acknowledgments

We are grateful to R. Hough for technical assistance with protein purification and critiques of this manuscript. We thank S. Beeser, N. King and the members of our laboratory for helpful discussions and R. Harland (University of California, Berkeley) for providing the *Xenopus* head library. Oligonucleotides were synthesized by the Utah Cancer Center Protein/DNA core facility supported by the National Cancer Institute (Grant PC30CA42014) or the Howard Hughes Medical Institute oligonucleotide synthesis facility at the University of Utah supported by the Department of Energy (Grant DE-FG03-94ER61817); cDNA-dsRBP-ZFa was sequenced by the University of Utah Health Sciences Sequencing Facility supported by the National Cancer Institute (Grant 5P30CA42014). This work was supported by funds to B. L. B. from the David and Lucile Packard Foundation. B. L. B. is an HHMI assistant investigator.

References

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

Andreazzoli, M., De-Lucchini, S., Costa, M. & Barsacchi, G. (1993). RNA binding properties and evolutionary conservation of the *Xenopus* multifinger protein Xfin. *Nucl. Acids. Res.* **21**, 4218–4225.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1990). Editors of *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences, New York.

Basham, B., Schroth, G. P. & Ho, P. S. (1995). An A-DNA triplet code: thermodynamic rules for predicting *A*- and *B*-DNA. *Proc. Natl Acad. Sci. USA,* **92**, 6464–6468.

Bass, B. L., Hurst, S. R. & Singer, J. D. (1994). Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr. Biol.* **4**, 301–314.

Berg, J. M. (1990). Zinc finger domains: hypotheses and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405–421.

Bevilacqua, P. C. & Cech, T. R. (1996). Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry,* **35**, 9983–9994.

Court, D. (1993). In *RNA Processing and Degradation by RNase III. Control of Messenger RNA Stability* (Belasco, J. G. & Brawerman, G., eds), Academic Press, Inc., San Diego.

Creighton, T. E. (1984). *Proteins.* 1st edit. W.H. Freeman and Co., New York.

Delarue, M. & Moras, D. (1989). RNA Structure. In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M. J., eds), vol. 3, pp. 182–196, Springer-Verlag, Heidelberg.

Denzler, K. L. & Jacobs, B. L. (1994). Site-directed mutagenic analysis of reovirus sigma3 protein binding to dsRNA. *Virology,* **204**, 190–199.

Dingwall, C. & Laskey, R. A. (1991). Nuclear targeting sequences: a consensus?. *Trends Biochem. Sci.* **16**, 478–481.

Doering, D. S. (1992). Functional and structural studies of a small F-actin binding domain. Ph D thesis, Massachusetts Institute of Technology, Boston, MA.

Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. (1992). Sequence-specific DNA binding by a two zinc-finger peptide from the *Drosophila melanogaster* Tramtrack protein. *J. Mol. Biol.* **226**, 349–366.

Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature,* **366**, 483–487.

Fasano, L., Roder, L., Core, N., Alexandre, E., Vola, C., Jacq, B. & Kerridge, S. (1991). The gene teashirt is required for the development of *Drosophila* embryonic trunk segments and encodes a protein with widely spaced zinc finger motifs. *Cell,* **64**, 63–79.

Geballe, A. P. & Morris, D. R. (1994). Initiation codons within 5′-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* **19**, 159–164.

Gibson, T. J. & Thompson, J. D. (1994). Detection of dsRNA-binding domains in RNA helicase A and *Drosophila* maleless: implications for monomeric RNA helicases. *Nucl. Acids Res.* **22**, 2552–2556.

Ginsberg, A. M., King, B. O. & Roeder, R. G. (1984). *Xenopus* 5 S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. *Cell,* **39**, 479–489.

Gonzalez, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M. & James, T. L. (1994). Structural study of a DNA.RNA hybrid duplex with a chiral phosphorothioate moiety by NMR: extraction of distance and torsion angle constraints and imino proton exchange rates. *Biochemistry,* **33**, 11062–11072.

Gurdon, J. B. & Wickens, M. P. (1983). The use of *Xenopus* oocytes for the expression of cloned genes. *Methods Enzymol.* **101**, 370–386.

Hemmati-Brivanlou, A., de la Torre, J. R., Holt, C. & Harland, R. M. (1991). Cephalic expression and molecular characterization of *Xenopus* En-2. *Development,* **111**, 715–724.

Hoovers, J. M., Mannens, M., John, R., Bliek, J., van Heyningen, V., Porteous, D. J., Leschot, N. J., Westerveld, A. & Little, P. F. (1992). High-resolution localization of 69 potential human zinc finger protein genes: a number are clustered. *Genomics,* **12**, 254–263.

Jacobs, G. H. (1992). Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J.* **11**, 4507–4517.

Jasanoff, A. & Weiss, M. A. (1993). Aromatic-histidine interactions in the zinc finger motif: structural inequivalence of phenylalanine and tyrosine in the hydrophobic core. *Biochemistry,* **32**, 1423–1432.

Joho, K. E., Darby, M. K., Crawford, E. T. & Brown, D. D. (1990). A finger protein structurally similar to TFIIIA that binds exclusively to 5 S RNA in *Xenopus*. *Cell,* **61**, 293–300.

Klug, A. & Schwabe, J. W. (1995). Zinc fingers. *FASEB J.* **9**, 597–604.

Koster, M., Kuhn, U., Bouwmeester, T., Nietfeld, W., el, B. T., Knochel, W. & Pieler, T. (1991). Structure, expression and in vitro functional characterization of a novel RNA binding zinc finger protein from *Xenopus*. *EMBO J.* **10**, 3087–3093.

Kozak, M. (1994). Determinants of translational fidelity and efficiency in vertebrate mRNAs. *Biochimie,* **76**, 815–821.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* **15**, 8125–8148.

LaCasse, E. C. & Lefebvre, Y. A. (1995). Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucl. Acids Res.* **23**, 1647–1656.

McGhee, J. D. & von Hippel, P. H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**, 469–489.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P. H. & Higuchi, M. (1996). A mammalian RNA editing enzyme. *Nature,* **379**, 460–464.

Mitchell, P. J. & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science,* **245**, 371–378.

Morinaga, T., Yasuda, H., Hashimoto, T., Higashio, K. & Tamaoki, T. (1991). A human alpha-fetoprotein enhancer-binding protein, ATBF1, contains four

homeodomains and seventeen zinc fingers. *Mol. Cell. Biol.* **11**, 6041–6049.
Nakai, K. & Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics,* **14**, 897–911.
Nelles, L. P. & Bamburg, J. R. (1976). Rapid visualization of protein-dodecyl sulfate complexes in polyacrylamide gels. *Anal. Biochem.* **73**, 522–531.
Nicholson, A. W. (1996). Structure, reactivity, and biology of double-stranded RNA. *Prog. Nucl. Acid Res.* **52**, 1–65.
Nightingale, K. P. & Wolffe, A. P. (1995). The interaction of TFIIIA with specific RNA-DNA heteroduplexes. *J. Biol. Chem.* **270**, 22665–22668.
Onate, A., Herrera, L., Antonelli, M., Birnbaumer, L. & Olate, J. (1992). *Xenopus* laevis oocyte G alpha subunits mRNAs. Detection and quantitation during oogenesis and early embryogenesis by competitive reverse PCR. *FEBS Letters,* **313**, 213–219.
Pavletich, N. P. & Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* **252**, 809–817.
Peabody, D. S. & Berg, P. (1986). Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Mol. Cell. Biol.* **6**, 2695–2703.
Pieler, T. & Bellefroid, E. (1994). Perspectives on zinc finger protein function and evolution–an update. *Mol. Biol. Reports,* **20**, 1–8.
Polson, A. G. & Bass, B. L. (1994). Preferential selection of adenosines for modification by double- stranded RNA adenosine deaminase. *EMBO J.* **13**, 5701–5711.
Polson, A. G., Bass, B. L. & Casey, J. L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature,* **380**, 454–456.
Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature,* **335**, 683–689.
Qian, X. Y., Chien, C. Y., Lu, Y., Montelione, G. T. & Krug, R. M. (1995). An amino-terminal polypeptide fragment of the influenza virus NS1 protein possesses specific RNA-binding activity and largely helical backbone structure. *RNA,* **1**, 948–956.
Qiu, Y., Nemeroff, M. & Krug, R. M. (1995). The influenza virus NS1 protein binds to a specific region in human U6 snRNA and inhibits U6-U2 and U6-U4 snRNA interactions during splicing. *RNA,* **1**, 304–316.
Reddy, B. A., Kloc, M. & Etkin, L. (1991). The cloning and characterization of a maternally expressed novel zinc finger nuclear phosphoprotein (xnf7) in *Xenopus* laevis. *Dev. Biol.* **148**, 107–116.
Saccomanno, L. & Bass, B. L. (1994). The cytoplasm of *Xenopus* oocytes contains a factor that protects double- stranded RNA from adenosine-to-inosine modification. *Mol. Cell. Biol.* **14**, 5425–5432.
Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S. & Reid, B. R. (1993). The DNA strand in DNA.RNA hybrid duplexes is neither *B*-form nor *A*-form in solution. *Biochemistry,* **32**, 4207–4215.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). In *Molecular Cloning: A Laboratory Manual* (Nolan, C., ed.), 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Scopes, R. K. (1987). In *Protein Purification Principles and Practice* (Cantor, C. K., ed.), 2nd edit., Springer Advanced Texts in Chemistry, Springer-Verlag, New York.
Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA,* **73**, 804–808.
Shi, Y. & Berg, J. M. (1995). Specific DNA-RNA hybrid binding by zinc finger proteins. *Science,* **268**, 282–284.
Singh, H., LeBowitz, J. H., Baldwin, A. S., Jr & Sharp, P. A. (1988). Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. *Cell,* **52**, 415–423.
St Johnston, D., Brown, N. H., Gall, J. G. & Jantsch, M. (1992). A conserved double-stranded RNA-binding domain. *Proc. Natl Acad. Sci. USA,* **89**, 10979–10983.
Steitz, T. A. (1993). Similarities and differences between RNA and DNA recognition by proteins. In *The RNA World* (Gesteland, R. F. & Atkins, J. F., eds), pp. 219–237, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Taylor, M. V., Gusse, M., Evan, G. I., Dathan, N. & Mechali, M. (1986). *Xenopus* myc proto-oncogene during development: expression as a stable maternal mRNA uncoupled from cell division. *EMBO J.* **5**, 3563–3570.
Theunissen, O., Rudt, F., Guddat, U., Mentzel, H. & Pieler, T. (1992). RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell,* **71**, 679–690.
Tsang, S. S., Yin, X., Guzzo-Arkuran, C., Jones, V. S. & Davison, A. J. (1993). Loss of resolution in gel electrophoresis of RNA: a problem associated with the presence of formaldehyde gradients. *Biotechniques,* **14**, 380–381.
Wickens, M. (1990). How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.* **15**, 277–281.
Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W. & Schreiber, S. L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell,* **76**, 933–945.

Edited by D. E. Draper

(Received 31 March 1997; received in revised form 30 May 1997; accepted 30 May 1997)