A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain

J. G. McAfee, L. Shahied-Milam, S. R. Soltaninassab and W. M. LeStourgeon

RNA 1996 2: 1139-1152

RNA (1996), 2:1139–1152. Cambridge University Press. Printed in the USA.

A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain

JAMES G. McAFEE,* LILLIAN SHAHIED-MILAM,* SYRUS R. SOLTANINASSAB, and WALLACE M. LeSTOURGEON

Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235, USA

ABSTRACT

The C protein tetramer of hnRNP complexes binds approximately 150–230 nt of RNA with high cooperativity (McAfee J et al., 1996, *Biochemistry 35*:1212–1222). Three contiguously bound tetramers fold 700-nt lengths of RNA into a 19S triangular intermediate that nucleates 40S hnRNP assembly in vitro (Huang M et al., 1994, *Mol Cell Biol 14*:518–533). Although it has been assumed that the consensus RNA recognition motif (RRM) of C protein (residues 8–87) is the primary determinant of RNA binding, we report here that a recombinant construct containing residues 1–115 has very low affinity for RNA at physiological ionic strength (100 mM NaCl). Moreover, we demonstrate that an N-terminal deletion construct lacking the consensus RRM but containing residues 140–290 binds RNA with an affinity sufficient to account for the total free energy change observed for the binding of intact protein. Like native C protein, the 140–290 construct is a tetramer in solution and binds RNA stoichiometrically in a salt-resistant manner in 100–300 mM NaCl. Residues 140–179 of the N-terminal truncated variant contain 11 basic and 2 acidic residues, whereas residues 180–207 specify a leucine zipper motif that directs dimer assembly. Elements within the 50-residue carboxy terminus of C protein are required for tetramer assembly. A basic region followed by a leucine zipper is identical to the domain organization of the basic-leucine zipper (bZIP) class of DNA binding proteins. Sequence homologies with other proteins containing RRMs and the bZIP motif suggest that residues 140–207 represent a conserved bZIP-like RNA binding motif (designated bZLM). The steric orientation of four high-affinity RNA binding sites about rigid leucine zipper domains may explain in part C protein's asymmetry, its large occluded site size, and its RNA folding activity.

Keywords: C protein; hnRNP; leucine zipper; ribonucleoprotein; RNA recognition motif

INTRODUCTION

In the presence of nuclease inhibitors, the great majority of the pre-mRNA molecules released from isolated HeLa S3 nuclei are recovered as poly hnRNP complexes that sediment from about 30 to 300S in density gradients (Beyer & Osheim, 1990; LeStourgeon et al., 1990; reviewed in McAfee et al., 1996a). Upon brief exposure to nuclease, these complexes are converted to 30–40S monoparticles that possess the same general protein composition as polyparticle complexes. The most abundant proteins (the "core" proteins A1, A2, B1, B2, C1, and C2) possess the intrinsic ability to spontaneously form 40S monoparticles on approximately 700-nt lengths of RNA and each multiple of this length supports the assembly of an additional particle (Conway et al., 1988; Huang et al., 1994). Based on hydrodynamic and protein crosslinking studies, it was suggested initially that most of the core proteins exist as tetramers of $(A1)_3B2$, $(A2)_3B1$, and $(C1)_3C2$ (Barnett et al., 1989, 1991). However, when expressed separately in bacterial cells, proteins C1 and C2 spontaneously form homo $C1_4$ and $C2_4$ tetramers (McAfee et al., 1996b). It is therefore possible that homotypic oligomers of the core proteins also exist in vivo.

In actively growing HeLa S3 cells, the nuclear concentration of C protein is about one-third to one-half that of histone H3 and is one of the three most abundant RNA binding proteins in the nucleus (LeStourgeon et al., 1981, 1990; Kiledjian et al., 1994). Hydrodynamic, ultrastructural, and equilibrium binding studies reveal that C protein's occluded RNA binding site size is between 150 and 230 nt (Huang et al., 1994; McAfee et al., 1996b). This large site size indicates that RNA may span

Reprint requests to: Wallace M. LeStourgeon, Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235, USA; e-mail: lestouwm@ctrvax.vanderbilt.edu.

*The first two authors contributed equally to the work.

a considerable distance between distally spaced RBDs in the tetramer. C protein binds RNA in a highly cooperative manner (McAfee et al., 1996b), and contiguous groupings of three tetramers undergo a combinatorial event to fold 700-nt increments of RNA into a unique triangular 19S complex (Huang et al., 1994; Rech et al., 1995a). Two triangular structures are observed when six tetramers are bound to 1,400 nt of RNA, whereas 2,100 nt support nine C protein tetramers forming a string of three triangles. The same RNA lengths support the in vitro assembly of hnRNP complexes composed of two or three 40S core particles, respectively (Conway et al., 1988; Huang et al., 1994). Additionally, purified 19S C protein–RNA complexes nucleate monoparticle assembly in vitro when added to dissociated preparations of crude hnRNP complexes, and they are recovered as remnant structures following dissociation of native hnRNP at 300 mM salt (Huang et al., 1994; Rech et al., 1995a). These findings indicate that, during transcription, C protein is likely to be the first hnRNP to associate with elongating pre-mRNA molecules and that the tetramer may function as an RNA chaperonin to constrain defined lengths of RNA in a state accessible by *trans*-acting factors (reviewed in Herschlag, 1995).

The C protein tetramer is very stable in solution. It retains its oligomeric structure in 2 M NaCl in the presence of β-mercaptoethanol, EDTA, and 0.5% deoxycholate, and it does not dissociate at pH 5.5 or 11.5 (Barnett et al., 1989). Based on the primary structure of C1 and C2, the molecular mass of the tetramer is 129 kDa (Swanson et al., 1987; Merrill et al., 1989) and it sediments in velocity sedimentation experiments at 5.8S with an apparent mass of 135 kDa. However, it elutes from size-exclusion columns as an asymmetric complex with an apparent mass of 390–420 kDa (Barnett et al., 1989; Huang et al., 1994). As reported here, it is likely that this anomaly is due to spatial asymmetry induced by the presence of two 30–35-Å coiled-coil domains or by the presence of a single four-coiled bundle. Additionally, it is likely that the coiled-coil domains orient the four salt-resistant RNA binding sites in a distally spaced manner.

The 90–100-residue amino-terminal domains of the C1 and C2 protomers constitute a consensus RNA recognition motif (RRM) or consensus RNA binding domain (RBD) (Keene & Query, 1991; Burd & Dreyfuss, 1994). Within this region, there exist two conserved octa- and hexa-peptide sequences denoted RNP-1 and RNP-2, respectively. The results of structural analyses by NMR and X-ray crystallography of several isolated RRMs show that they assume a common folding pattern consisting of a $\beta1\alpha1\beta2\beta3\alpha2\beta4$ secondary structure (Nagai et al., 1990; Hoffman et al., 1991; Wittekind et al., 1992; Lee et al., 1994; Lu & Hall, 1995). More specifically, a 10-nt loop of U1 snRNA binds the β-sheet of the U1A protein as an open structure, and a specific loop sequence interacts with the consensus RNP1 and 2 motifs as well as the C-terminal extension of the RNP domain (Oubridge et al., 1994). As reported here, however, C protein constructs lacking the canonical RRM retain the unique RNA binding properties of the native protein (high affinity under equilibrium conditions, salt-resistant binding, and specificity for a SELEX-identified "winner" sequence). These RNA binding activities are dependent on a region of basic residues preceding a leucine zipper domain that drives dimer formation and tetramer assembly in the presence of the 50-residue carboxy-terminal domain.

RESULTS

Identification of residues involved in C protein oligomerization

Multidimensional NMR studies on the isolated RRM of C protein (residues 1–94) (Wittekind et al., 1992) and equilibrium binding studies on a similar fragment (9–102) (Amrute et al., 1994) have shown that the amino-terminal domain is monomeric in solution. Thus, residues carboxy-terminal to Ala 102 are involved in the oligomerization of C protein. To identify the primary determinants involved in tetramer assembly, a series of carboxy- and amino-terminal truncations of protein C1 were generated, subcloned in *Escherichia coli*, expressed, and purified using Ni-histidine columns (Fig. 1). Velocity sedimentation analysis and size-exclusion chromatography were utilized to analyze the oligomeric state of the deletion constructs. The full-length fusion protein (Met 1–Gly 290) was used as a control to insure that the amino-terminal His tag does not interfere with oligomerization or the sedimentation properties of the tetramer. The sedimentation coefficient of the His tag protein was indistinguishable from that of native and recombinant $C1_4$ and $C2_4$ tetramers (5.8S) (not shown). The Met 1–Phe 115 construct sediments as a monomer with a calculated molecular mass of 13 kDa (1.5S) (Fig. 2A). Likewise, the Met 1–Leu 180 construct is also monomeric and sediments with an apparent mass of 29 kDa (2.5S) (Fig. 2B). These results were confirmed by the observation that both constructs elute from size-exclusion columns as monomers (not shown). Shown in Figure 2C are the results of sedimentation experiments using the Met 1–Ser 240 construct. The calculated sedimentation coefficient and molecular mass of the Met 1–Ser 240 mutant are, respectively, 4.3S and 66 kDa. These values correspond to that expected for a dimer of this construct (57,596 Da with His tag) and indicate that residues Leu 180–Ser 240 are involved in the dimerization of C protein. Because both native and recombinant C protein are stable tetramers in solution, these results further indicate that regions within the 50-residue carboxy terminus (Ser 241–Gly 290) are re-

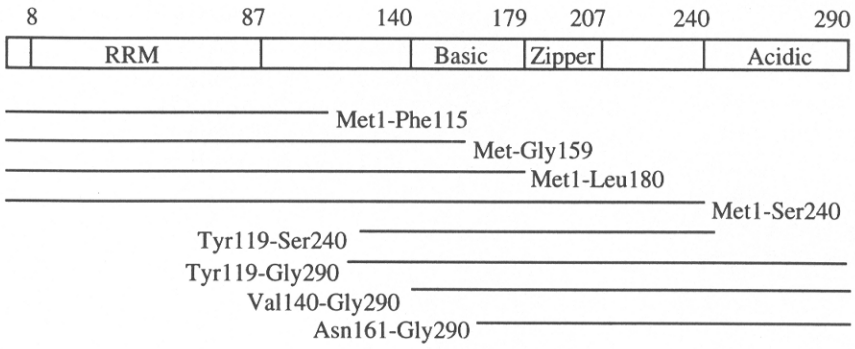


FIGURE 1. Primary structural motifs and deletion constructs of hnRNP C1. Shown at the top is a schematic diagram illustrating the position of functional regions in the primary structure of hnRNP C1, including the RNA recognition motif (RRM); a highly basic region involved in RNA binding and extending from Val 140 to Asp 179 (Basic); the four heptad consensus repeat, specifying a leucine zipper (Zipper); and a region of high acidity (Acidic). Below the diagram are shown the amino- and carboxy-terminal boundaries of deletion constructs used in this study. Each construct is denoted by the three-letter abbreviation for the terminal residues as shown for the published sequence of hnRNP C1 (Swanson et al., 1987).

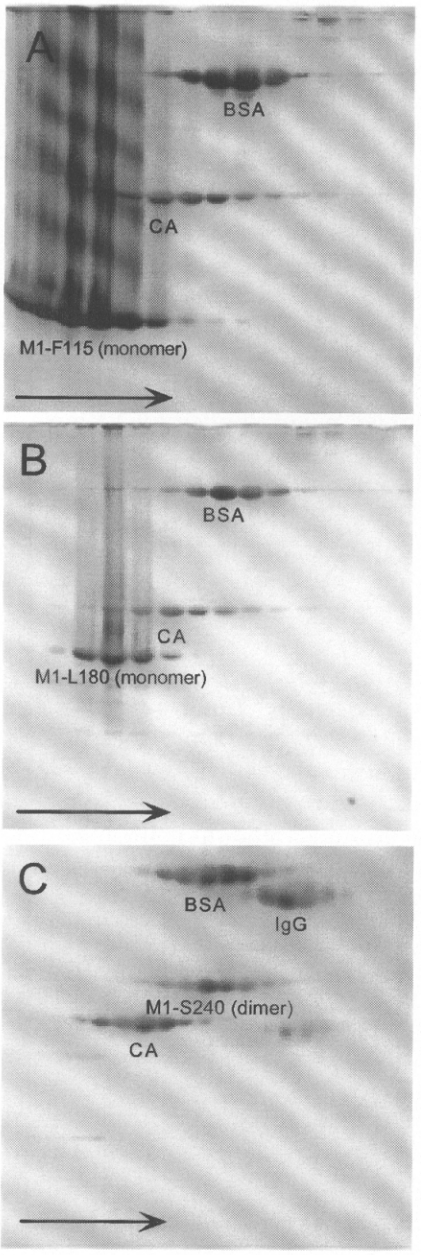
quired for tetramer stability and that the tetramer may be a dimer of dimers.

C protein contains a coiled-coil domain:

An examination of the Met 1–Ser 240 dimerizing construct revealed the presence of a consensus coiled-coil or "leucine zipper" domain between residues Leu 180 and Glu 207 (Fig. 3A). These domains are characterized by the presence of apolar residues at every seventh position and also usually at the fourth position. Residues within each heptad repeat are designated a, b, c, d, e, f, g (Crick, 1953; Cohen & Parry, 1986, 1994). The distinctive features of coiled-coil domains within fibrous proteins are: (1) a predominance of Leu, Ile, Val, Met, Phe, and Tyr in the a and d positions; (2) approximately 50% of the residues at the b, c, e, f, and g positions are charged; (3) the charged to apolar residue ratio is between 0.9 and 1.4; and (4) the sequence has a high propensity to form an amphipathic α-helix, where the inner core of the coiled-coil is characterized by hydrophobic interactions, whereas the outer ionic interactions serve to stabilize the structure (Cohen & Parry, 1990, 1994). In C protein, the residues between Leu 180 and Glu 207 are characteristic of classical coiled-coil domains. Apolar amino acids, most of which are hydrophobic, are located in the (a) and (d) positions, whereas many charged residues are found at the other sites (b, c, e, f, g). The charged to apolar ratio is 1.3 and the Chou–Fasman secondary structure prediction algorithm indicates that residues 180–214 have a high propensity to be α-helical. The calculated length of this helical region is at least 30 Å. Helical-wheel analysis of the C protein heptads reveal that a helical conformation of residues Leu 180–Glu 207 would position the apolar residues (a and d) on one side of the helix, whereas the charged residues are on the other side, as expected in coiled-coil domains (Fig. 3B). Additionally, when the 28-residue sequence (Leu 180–Glu 207) of C protein is analyzed with the PARICOIL algorithm (a new method of distinguishing paired coiled-coil domains from α-helices) (Berger et al., 1995), a correlation score of 0.84 is obtained. Scores of 0.5 and above are considered positive for coiled-coil regions in proteins.

To determine experimentally if C protein's coiled-coil domain possesses the intrinsic property to oligomerize under physiological conditions, CD performed on the Leu 180–Glu 207 peptide (see the Materials and methods) was utilized to distinguish between the presence of monomeric α-helix and coiled-coil structures. The spectra of the peptide at 7.4 mM concentration (residues) shows that the two minima at 208 nm and 222 nm have molar ellipticity values of $-27{,}827$ and $-27{,}473$ deg cm^2 dmol^{-1}, respectively, indicating that the peptide is 83% α-helical (assuming that $-33{,}000$ deg cm^2 dmol^{-1} is 100% α-helical, as described in Chen et al. [1974]) (Chen et al., 1974; O'Shea et al., 1989a; Zhou et al., 1992; Lumb et al., 1994) (Fig. 4A). To distinguish between monomeric α-helices and coiled-coil structures, the α-helix-inducing solvent trifluoroethanol (TFE) was added to the peptide at a 1:1 (v/v) ratio. As shown previously, TFE has a marginal effect on helices stabilized by coiled-coil interactions (Zhou et al., 1992), but dramatically increases the α-helical content of isolated helical peptides that exist in equilibrium between helical and random coil states (Saudek et al., 1991). As shown in Figure 4B, the peptide (3.7 mM residue) was affected minimally by the addition of TFE. These findings indicate that the Leu 180–Glu 207 peptide (the leucine zipper region of C protein) exists predominantly as a coiled-coil structure in physiological solution. Consistent with this evidence are preliminary thermal stability studies yielding a T_m of >89 °C for the Leu 180–Glu 207 peptide (not shown). As shown by Harbury et al. (1993), peptides possessing four heptad repeats may, depending on their residue composition, form dimeric, trimeric, or tetrameric coiled coils in solution. Interestingly, all of these oligomeric complexes can possess T_m values >100 °C.

To further demonstrate that residues 180–207 function in oligomerization, site-specific mutagenesis was performed to create a C1 mutant in which Leu 187 and Val 194 of the zipper motif were converted to alanines. The objective was to demonstrate the monomeric state of this protomer. The mutant was cloned into a pET-28a vector (Novagen) and transformed into a Bl21 DE3 (pLysS) strain of *E. coli* (see the Materials and methods) and the mutations were confirmed by sequencing. However, analysis via SDS-PAGE of total bacterial protein revealed negligible levels of the construct following IPTG induction. As described in the Discussion, this finding exists as indirect evidence that residues 180–207 play a critical role in C protein oligomerization, because incorrectly folded polypeptides are usually degraded rapidly in *E. coli*.



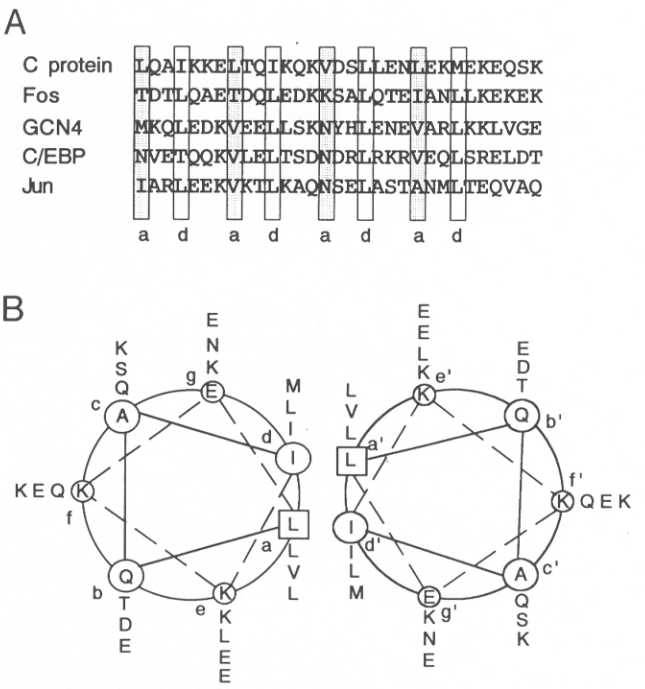


FIGURE 3. **A:** Homology between the leucine zipper region of C1 and representative members of the bZIP class of DNA binding proteins. The four heptad repeat of C1 is aligned with homologous regions of the transcription factors Fos, GCN4, C/EBP, and Jun. The apolar residues (primarily leucines) in the (a) position of each sequence are shaded. As shown, each protein contains four heptad repeats. **B:** Helical-wheel representation of the leucine zipper region of C protein. Shown are two helical wheels of parallel dimers of C1 corresponding to residues 180–207. Amino acids within this region are represented by single capital letters. Lower-case a–g or a′–g′ indicate the position of each amino acid in a single heptad for each monomer of the dimer.

A basic region preceding the coiled-coil domain is responsible for C proteins unique high-affinity and salt-resistant RNA binding activity

The carboxy-terminal deletion construct Met 1–Gly 159 exists in solution as a monomer and contains the amino-

FIGURE 2. Velocity sedimentation analysis of C protein deletion constructs. The deletional variants, Met 1–Phe 115 (**A**), Met 1–Leu 180 (**B**), and Met 1–Ser 240 (**C**) were co-sedimented with BSA (66 kDa), carbonic anhydrase (29 kDa), or IgG (150 kDa) as indicated in the Materials and methods. Shown are Coomassie blue-stained SDS-PAGE gels of successive fractions from each 5–20% sucrose gradient. Sedimentation coefficients were calculated for each construct according to the method of Martin and Ames (1961). Calculated sedimentation coefficients for Met 1–Phe 115, Met 1–Leu 180, and Met 1–Ser 240 were 1.5S, 2.5S, and 4.3S, respectively.

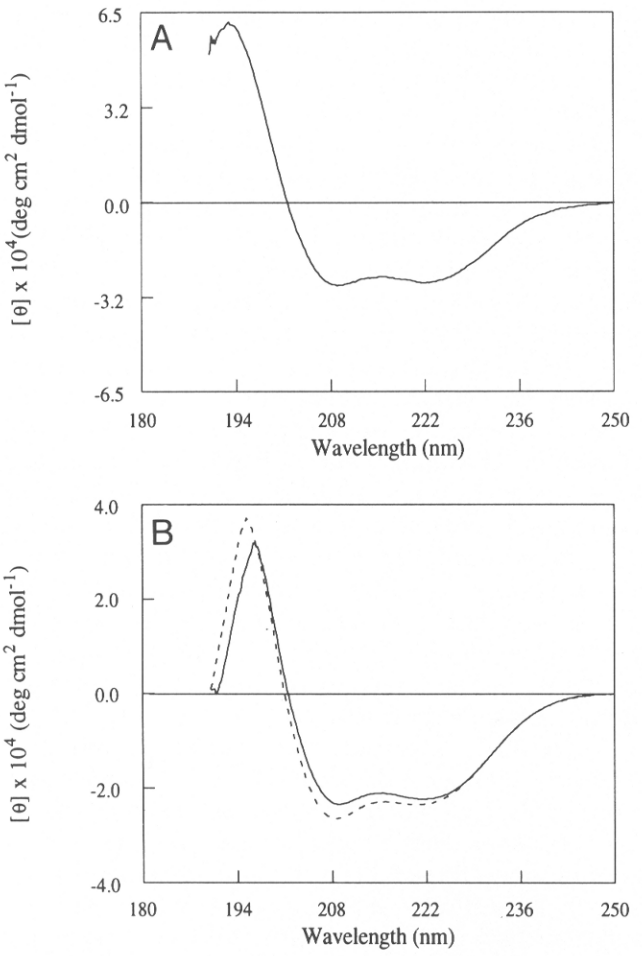


FIGURE 4. CD spectra of the Leu 180–Glu 207 peptide under oxidizing conditions. **A:** Normalized spectra of a 7.4-mM (residues) sample at room temperature. The $[\theta]_{208}$ and $[\theta]_{222}$ are −27,827 and −27,473 deg cm^2 dmol^{-1}, respectively. These values are characteristic of α-helical species. **B:** Spectra of the peptides (3.7-mM residues) in a 1:1 (v/v) ratio with TFE (dashed line) shows little variance from the peptide in the absence of TFE (continuous line), indicating that this α-helix-inducing solvent has a marginal effect on the structure of the peptide.

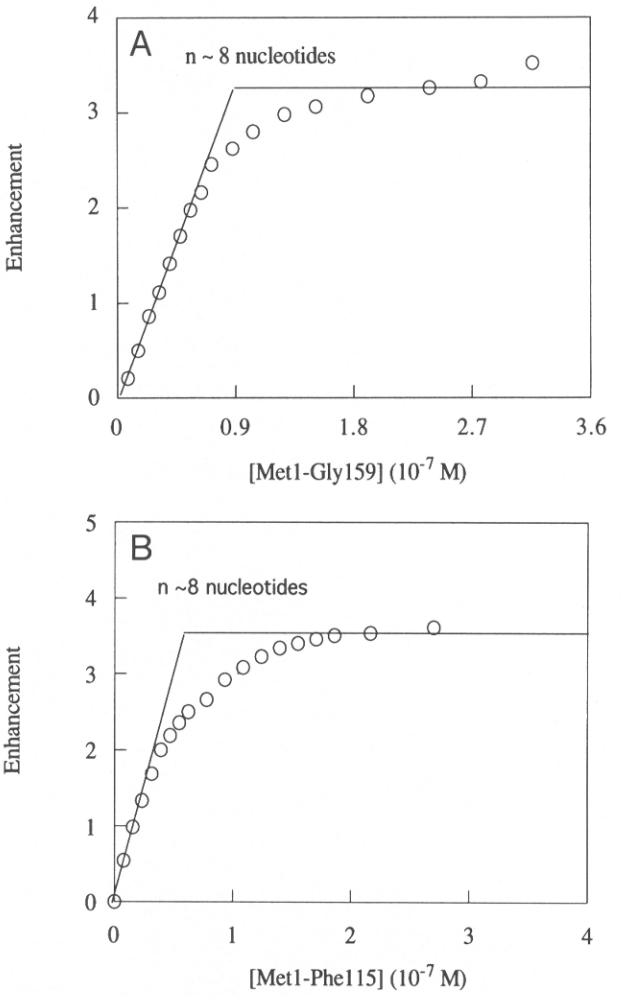


FIGURE 5. Stoichiometry of the Met 1–Gly 159 and the Met 1–Phe 115 constructs binding to poly[r(ϵA)]. A 7×10^{-7} M, or an 8×10^{-7} M solution of poly[r(ϵA)] was titrated with the Met 1–Gly 159 construct (**A**) or with the Met 1–Phe 115 construct (**B**), respectively. Each titration was performed in the absence of NaCl to facilitate stoichiometric binding. Protein concentrations are expressed in molar polypeptides. The RNA/protein ratio (occluded binding site size) at RNA saturation is estimated from both graphs by the intersection of the two straight lines in each panel. The site size (n) is indicated in each panel.

terminal consensus RRM that spans residues 8–87 (Wittekind et al., 1992; Gorlach et al., 1994). Previous equilibrium binding studies conducted on a proteolytic fragment of C protein (residues 9–102) yielded an occluded RNA binding-site size of ~7 nt (Amrute et al., 1994). To determine the occluded site size for the Met 1–Gly 159 construct, a titration was conducted in the absence of NaCl to facilitate stoichiometric binding. As shown in Figure 5A, the RNA/protein ratio at RNA saturation corresponds to a binding site size of ~8 nt. At approximately physiological ionic strength (100 mM NaCl), the polypeptide binds polyethenoadenylic acid (poly[r(ϵA)]) with an apparent affinity of 2×10^7 M^{-1}. Thus, whereas the site size observed for Met 1–Gly 159 agrees with that reported previously for the 9–102 C protein fragment, its affinity at 100 mM salt is ~100-fold higher than that reported previously for the 9–102 fragment (3×10^5 M^{-1}). Additionally, in the absence of salt, a Met 1–Phe 115 construct also binds RNA with an occluded site size of ~8 nt (Fig. 5B), but, at 100 mM salt, this construct binds RNA with exceptionally low affinity. These initial findings indicated that residues between 115 and 159 dramatically increase the RNA binding affinity of the Met 1–Gly 159 polypeptide.

To further define the residues responsible for C protein's high-affinity and salt-resistant RNA binding activity, and to explore the role of oligomerization on these activities, we characterized the binding properties of several amino-terminal truncation variants. As shown in Figure 6A, the almost linear shape of the ascending part of the binding isotherm and the sharp transition at maximum enhancement demonstrates

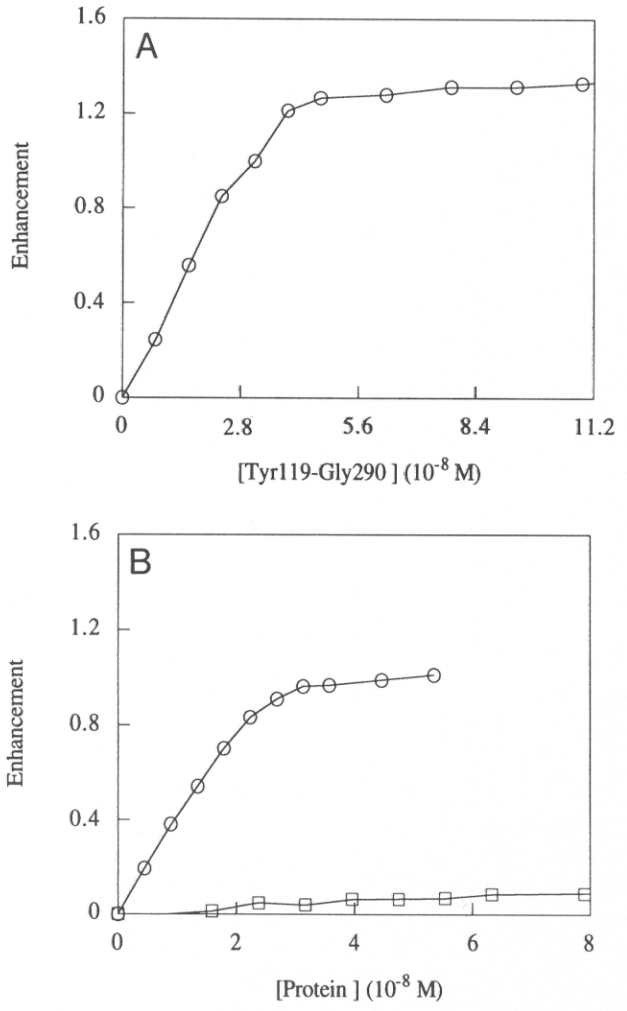


FIGURE 6. Identification of the carboxy-terminal binding determinants of hnRNP C1. An 8×10^{-7} M solution of poly[r(ϵA)] was titrated with the Tyr 119–Gly 290 construct (**A**), the Val 140–Gly 290 construct (**B**, circles), or the Asn 161–Gly 290 construct (**B**, squares). Lines through the data in both panels are interpolated from the data points.

that the Tyr 119–Gly 290 construct binds RNA with high affinity at 100 mM NaCl. This construct contains the coiled-coil motif and the carboxy-terminal elements responsible for tetramerization, and sediments as a tetramer in sucrose gradients (not shown). An extrapolation to stoichiometric binding yields an occluded site size of approximately 160 nt. Although wild-type C1 and the Tyr 119–Gly 290 construct have comparable occluded binding-site sizes, the affinity of the latter at 100 mM NaCl (binds stoichiometrically) exceeds that of the wild-type protein (i.e., the association constant for $C1_4$ is approximately 3×10^6 M^{-1}) (McAfee et al., 1996b). To determine if the 50-residue carboxy-terminal region of C protein contributes significantly to RNA binding activity and to determine if tetramerization is required for C protein's high binding affinity, the binding properties of the dimeric deletion construct (Tyr 119–Ser 240) were determined at 100 mM NaCl. The binding isotherms of the Tyr 119–Gly 290 and the Tyr 119–Ser 240 constructs were essentially identical (not shown).

The above findings indicate that C protein's high-affinity and salt-resistant RNA binding activity lies outside the canonical RRM and is not dependent on tetramer assembly or the 50-residue carboxy-terminal region. To identify the specific region involved in RNA binding, we sequentially removed residues downstream of Tyr 119 to generate the Val 140–Gly 290 and the Asn 161–Gly 290 constructs. Both of these deletion constructs contain the coiled-coil motif and the carboxy-terminal region of C protein and exist as tetramers in solution. Shown in Figure 6B are binding isotherms for the interaction of Val 140–Gly 290 and Asn 161–Gly 290 with poly[r(ϵA)] in 100 mM NaCl. The Asn 161–Gly 290 construct shows little or no RNA binding activity, whereas the interaction of the Val 140–Gly 290 variant is qualitatively similar to that observed for the Tyr 119–Gly 290 protein. Thus, residues between Val 140 and Asn 161 are critical for C protein's unique RNA binding activity.

Analysis of the primary sequence of C protein reveals an asymmetric distribution of charged residues. Specifically, a high percentage of basic residues are located between Tyr 119 and Ser 240. Binding studies conducted on the wild-type C1 and C2 proteins indicate that the affinity of the protein is not affected significantly between 100 and 300 mM NaCl (not shown). This indicates that the electrostatic contribution to the free energy of binding is relatively small over this salt range. However, it is possible that removing the amino-terminal portion of the protein generates a nonspecific electrostatic binding component that does not function in the intact protein. To test this, we examined the salt sensitivity of the binding of the Tyr 119–Gly 290 construct to determine if it exhibits the same behavior as the wild-type protein. Shown in Figure 7A are binding isotherms for the interaction of Tyr 119–Gly 290 with poly[r(ϵA)] at 100, 200, and 300 mM NaCl. Like native and recombinant C protein, the quantum yield of the Tyr 119–Gly 290–poly[r(ϵA)] complex decreases with increasing ionic strength. This suggests that both native C protein and the Tyr 119–Gly 290 construct undergo similar conformational changes upon increasing ionic strength. However, at elevated ionic strength, the RNA saturation transition for each binding isotherm occurs at approximately the same protein concentration, indicating little difference in affinity of the protein at 100 and 200 mM NaCl. The affinities of recombinant C1, C2, and native protein also show little change in this salt range, further indicating that intact C protein and the Tyr 119–Gly 290 construct bind RNA through similar mechanisms (McAfee et al., 1996b). Additional evidence that residues carboxy-terminal to Val 140 are essential for binding is seen in the salt-resistant bind-

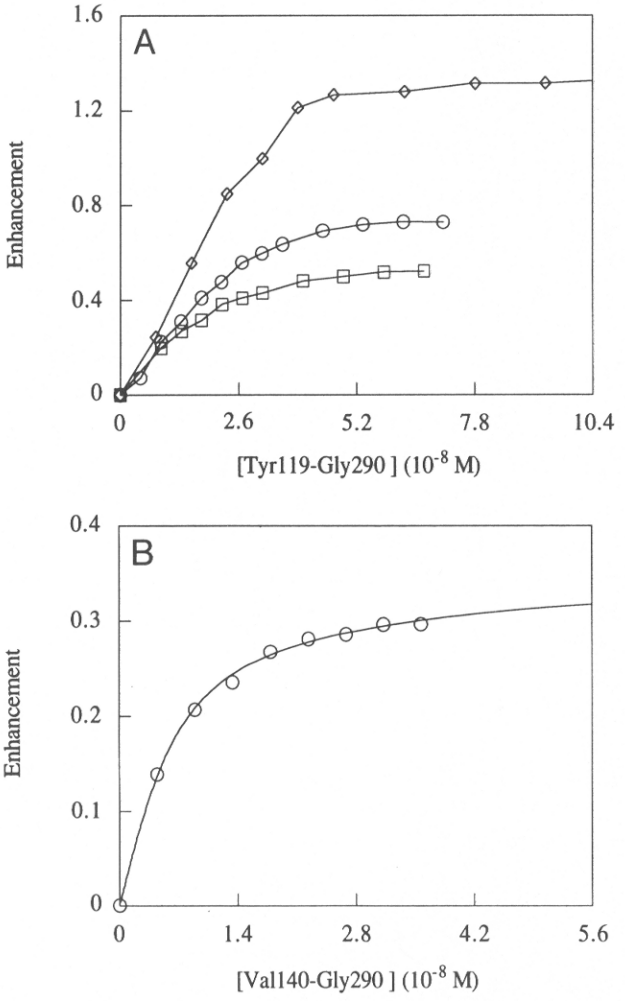


FIGURE 7. Effect of ionic strength on the RNA binding affinity of the Tyr 119–Gly 290 and the Val 140–Gly 290 constructs. An 8×10^{-7} M solution of poly[r(ϵA)] was titrated with the Tyr 119–Gly 290 construct in the presence of 100 mM (diamonds), 200 mM (circles), or 300 mM (squares) NaCl. **A:** Lines are interpolated from the data points. **B:** An 8×10^{-7} M solution of poly[r(ϵA)] containing 300 mM NaCl was titrated with the Val 140–Gly 290 construct. The continuous line through the data points was simulated using the McGhee–von Hippel noncooperative binding equation assuming an association constant of 5.64×10^6 M^{-1} and a binding site size of ~75 nt.

ing of the Val 140–Gly 290 construct in Figure 7B. The McGhee–von Hippel (1974) noncooperative model was used to simulate the theoretical curve shown in the figure, assuming an association constant of 5.64×10^6 M^{-1} and a site size of 72 nt/monomer. These observations suggest that, like the full-length recombinant C1 and native C protein, the electrostatic contribution to the free energy of binding of the amino-terminal deletion variant is quite small at physiological ionic strength. This would not be expected if deletions create an artifactual electrostatic binding component. These findings also suggest that either hydrogen bonding, van der Waals, or hydrophobic interactions contribute a large percentage to the overall free energy of binding. This is identical to observations with recombinant and native C proteins, where the electrostatic contribution to the free energy of binding is less than 30% at 100 mM NaCl (unpubl. obs.).

Further evidence that the structural context of the RNA binding activity in native C protein is maintained in the Tyr 119–Gly 290 construct is found in the emission spectra of these proteins. More specifically, emission spectra of the Tyr 119–Gly 290 construct excited at 280 nm reveal the presence of a unique tyrosinate fluorophore (Ross et al., 1992). Because native and recombinant C protein possess this same unique emission spectrum (McAfee et al., 1996b), the conformational environment of the tyrosines undergoing this emission (Tyr 119 or Tyr 120) in native C protein are preserved in the deletion variant.

Basic high-affinity RNA binding region binds the SELEX-identified high-affinity substrate for C protein

Recently, recombinant C1 has been used to selectively isolate oligonucleotide substrates with high binding affinity from random pools of RNA (the SELEX procedure) (Gorlach et al., 1994). These studies defined high-affinity "winner sequences" for recombinant C1, which consist of five contiguous uridines. To determine if the RNA binding determinants that select the winner sequences reside in residues V140–G290, we performed competition binding experiments under equilibrium conditions with one of the SELEX winners (AGUAU UUUUGUGGA), and with a version randomized at the position of the five contiguous uridines (AGUAU CCAAGUGGA). As shown in Figure 8, the V140–G290 polypeptide binds the winner sequence and the randomized version of the winner with apparent affinities (K_{app}) 2.5×10^6 M^{-1} and 8.2×10^4 M^{-1}, respectively. The 30-fold higher affinity for the winner sequence further indicates that the major determinants of C protein's RNA binding properties reside outside of the canonical RRM and that the truncated variant binds RNA fundamentally like the wild-type protein.

Effect of oligomerization on RNA binding site size

Both native and recombinant C protein bind RNA with an occluded site size of 150–230 nt (Huang et al., 1994; McAfee et al., 1996b). Although electron micrographs of partially dehydrated C protein reveal a tetramer diameter of 85–90 Å (Huang et al., 1994; Rech et al., 1995b), the tetramer elutes from size-exclusion columns as an asymmetric complex with at least one diameter of 135 Å (Barnett et al., 1989; Huang et al., 1994). It is therefore possible that RNA wraps about the surface of the tetramer, making multiple contacts with distally spaced binding domains, or folds in some manner between distally spaced binding domains within the tet-

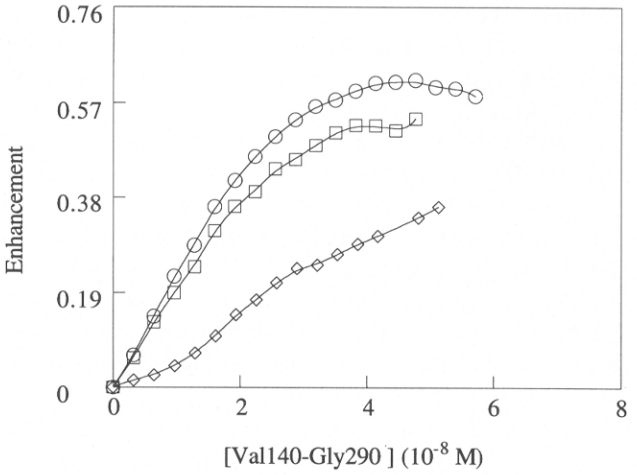


FIGURE 8. Competition binding experiments with a SELEX winner sequence, and a randomized version of this sequence. Solutions containing 8×10^{-7} M of poly[r(ϵA)] were titrated with the Val 140–Gly 290 construct in the absence of competitor (open circles), in the presence of a fivefold molar excess of the SELEX winner sequence (AGUAUUUUUGUGGA) (diamonds), or in the presence of a 10-fold molar excess of a sequence randomized at the position of the five contiguous uridines (AGUAUCCAAGUGGA) (squares). The apparent association constants for the winner and randomized version are 2.5×10^6 M^{-1} and 8.2×10^4 M^{-1}, respectively.

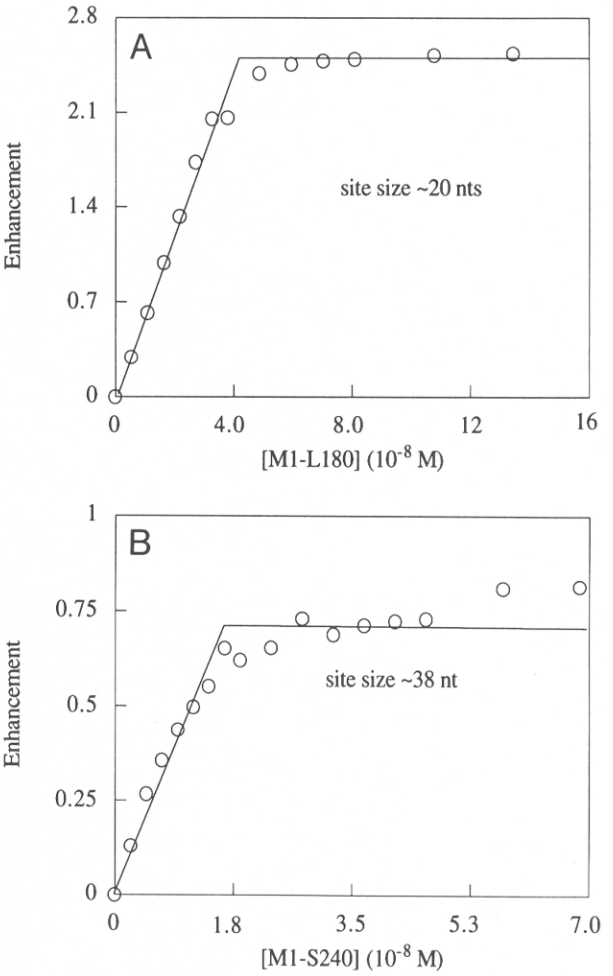


FIGURE 9. Determination of the occluded binding site size for the Met 1–Leu 180 and the Met 1–Ser 240 constructs. Shown are plots of the fluorescence enhancement observed for a 0.8 μM solution of poly[r(ϵA)] titrated with the Met 1–Leu 180 (**A**) or the Met 1–Ser 240 (**B**) constructs in buffer containing 10 mM Tris, pH 8.0, 1 mM EDTA, and 100 mM NaCl. In these experiments, the fractional change in the fluorescence enhancement is linearly correlated with the fractional saturation of the poly[r(ϵA)]. The linear ascending portion of each binding isotherm indicates that each variant binds poly[r(ϵA)] under these buffer conditions stoichiometrically. The RNA/protein ratio at poly[r(ϵA)] saturation (occluded binding site size) was determined from the intersection of the two straight lines shown in each panel. The binding site size for the Met 1–Leu 180 construct was calculated to be ~20 nt, whereas that of the Met 1–Ser 240 dimer was ~38 nt.

ramer. As shown above, the monomeric Met 1–Gly 159 construct binds RNA with an occluded site size near 8 nt. This site size is essentially the same as the amino-terminal RRM alone (Amrute et al., 1994). Consistent with a minimal RNA binding role for the RRM and with distally spaced high-affinity binding sites in the tetramer is the finding that the tetrameric construct Tyr 119–Gly 290 (lacking the amino-terminal RRM) binds RNA stoichiometrically with a site size of approximately 160 nt (Fig. 3A, described above). To gain further insight into the topology of the RBDs in the tetramer, we have determined the occluded RNA binding site sizes of the Met 1–Leu 180 monomer, which terminates at the beginning of the coiled-coil domain, and the Met 1–Ser 240 dimer, which terminates 33 residues C-terminal to the coiled-coil domain. As shown in Figure 9, the RNA/protein ratios at saturation correspond to binding-site sizes of 20 and 38 nt, respectively, for the monomeric and dimeric species. This indicates that the high-affinity RNA binding sites may be closely positioned in the dimer, but distally spaced in the tetramer, because the occluded site size for the latter is near 200 nt.

DISCUSSION

As shown above, C protein constructs that lack the 50-residue carboxy terminus (Glu 241–Met 290) exist as dimers, and those that lack the consensus coiled-coil domain (Leu 180–Glu 207) exist as monomers. As shown here, the Leu 180–Glu 207 peptide possesses the intrinsic property to form coiled-coil structures under physiological conditions. These findings, together with the well-characterized property of coiled-coil domains to oligomerize (O'Shea et al., 1989b; Harbury et al., 1993), indicates that the coiled-coil domain of C protein (Leu 180–Glu 207) is probably the major determinant of dimer assembly. Consistent with this activity are the primary sequence analyses that reveal that Leu 180–Glu 207 is a consensus leucine zipper domain. Within these 28 residues there are four heptad repeats of hydrophobic residues that can form a hydrophobic track along one surface of an alpha-helical mo-

tif, whereas ionic residues may interact through salt bridges between subunits (Cohen & Parry, 1986). Like the leucine zippers of well-characterized DNA-binding proteins (GCN4, Fos, C/EBP, and Jun) (Kouzarides & Ziff, 1988; O'Shea et al., 1989b; Schuermann et al., 1989), the coiled-coil of C protein has apolar residues in the a and d positions of the a–g heptad repeat. This further contributes to the amphipathic nature of C protein's coiled domain and to the hydrophobic interactions that stabilize their association. Interestingly, in C protein, leucine occupies the "a" or start position of three of the four heptads, whereas in the dimeric DNA binding proteins, leucine occupies the "d" or fourth position within each heptad repeat. Further evidence that C protein possesses a rod-like domain is the charged to apolar ratio (1.3) of the leucine zipper region. This value lies within the range for the coiled-coil domain of other fibrous proteins of 0.9–1.4 (Cohen & Parry, 1986). The presence of rod-like domains in C protein is also consistent with, and may explain, the protein's asymmetry and anomalously large apparent mass (about 390 kDa versus 130 kDa), as determined from size-exclusion chromatography (Barnett et al., 1989; Huang et al., 1994). It is also likely to explain its large RNA binding site size (about 230 nt).

The insignificant expression in bacterial cells of a construct containing alanines at critical sites within the zipper domain (Leu 187 and Val 194) is likely due to the exposure of protease recognition sites in an abnormally folded monomeric protein (Gottesman & Maurizi, 1992). For example, it has been shown that single amino acid changes in the oligomerization domain of the transcriptional activator λ *c*II result in monomeric variants (opposed to the native tetrameric configuration) with half-lives as low as 36 s. Also, monomeric mutants of the *Lac* repressor (also a tetramer) have very low expression in bacterial cells (Platt et al., 1970; Ho et al., 1988). These findings are not surprising because it is well established that proteins folded abnormally are usually degraded rapidly in *E. coli* (Goldberg & Dice, 1974; Goldberg & St. John, 1976; Gottesman & Maurizi, 1992). Finally, this phenomenon has been demonstrated in eukaryotic systems with the rapid degradation of oligomerization-defective hemoglobin peptides (Goldberg & St. John, 1976). The efficient expression of the monomeric carboxy-terminal deletion mutants of C protein lacking the leucine zipper motif (M1–F115 and M1–L180) further suggests that exposed protein–protein interaction domains in monomeric subunits are targeted by proteases.

Experiments performed using synthetic peptides of the leucine zipper regions of GCN4, Fos, Jun, and Lac repressor have shown that four heptad repeats are sufficient for stable dimer formation (O'Shea et al., 1991, 1989b; Chen et al., 1994). In these DNA binding proteins, the α-helical coil serves as a scaffolding for the two basic "arms," which extend to interact with the DNA at dyad-symmetric sites (Vinson et al., 1989). Homology searches of crystal structure databases selected the leucine zipper of GCN4 as the homologue of C protein's four heptad repeats (Leu 180–Glu 207). Like the dimeric transcription factors, the consensus leucine zipper domain of C protein immediately follows the basic salt-resistant RNA binding motif. Thus, fibrous coiled-coil domains may function to orient C protein's high-affinity RNA binding sites.

Regarding tetramer assembly, residues in the highly acidic carboxy terminus may associate with charged sites on the zipper domain and facilitate tetramer assembly through a "domain swapping" event (Bennett et al., 1995). In other words, the dimeric coiled-coil domains may associate to form a four-coiled bundle with all parallel or with alternating orientations. It is also possible that the tetramer possesses two separate leucine zipper domains and that the amino-terminal RBDs are distally spaced in an asymmetric tetrameric complex stabilized by interactions of the carboxy-terminal domains. This would be consistent with C protein's large RNA binding-site size. If the coiled domains associate in an antiparallel manner, the high-affinity sites would be distally spaced, at least at their carboxy ends, by about 30 Å. The separation of two nucleic acid binding domains by a 28-residue helix has been observed previously in histones H3 and H4 (Arents et al., 1991). The $H3_2H4_2$ tetramer is about one-third smaller than the C protein tetramer, but orders 146 bp of DNA about its surface. The topology of C protein's RBDs underlies the tetramer's unique ability to measure and bind relatively long lengths of RNA and constrain the RNA such that other protein binding events can occur during mRNA maturation.

The results reported here also demonstrate that the salt-resistant RNA binding activity of C protein is dependent on residues between Val 140 and Ser 240. Specifically, the removal of residues 140–160 abrogates binding of the protein to RNA, suggesting that critical determinants for binding are in this region. The carboxy-terminal boundary of the RNA binding activity in C protein was not mapped by deletional analysis in an effort to maintain the native quaternary structure of the protein. As described here, the region directly preceding the leucine zipper (Val 140–Glu 179) contains 11 basic and 2 acidic residues. A basic region amino-terminal to a leucine zipper is characteristic of the leucine zipper class of DNA binding proteins (Landschulz et al., 1988). In these proteins, it is the basic region that interacts with the DNA substrate. By analogy to this group of DNA binding proteins, we predict that the primary binding determinants of C protein reside in the 40-residue basic region (Val 140–Asp 179) preceding the leucine zipper. Homology searches of protein databases using this 40-residue sequence select core hnRNP A1, as well as other known and suspected RNA binding proteins and some DNA binding proteins (Fig. 10).

```
C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSG     KLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
      :: :: ::**    ::*:* *  : * : *  *       :::  *:::::  ::::::* *  :: *:: *  *
p542  VPVKRPRVTVPLVRRVKTNVPVKLFARSTAVTTSSAKIKLKSSELQAIKTELTQIKSNIDALLSRLEQIAAD
      ::::::::::::::::::**:::::::::::::: :::::::::::: ::::::::::::::: :::::: *
Merc  VPVKRPRVTVPLVRRVKTTIPVKLFARSTAVTTGSAKIKLKSSELQTIKTELTQIKSNIDALLGRLEQIAEE
      :: :: ::**    ::*:* *  : * : *  *       :::  *:: :: ::::::* *: :: *:: *  :
C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSG     KLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
         ::** *:***:*::*::  *  : ::    :
A1       SKQEMASASSSQRGRSGSGNFGGGRGGGFGG

C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSGKLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
                :    ::* ::  * : :: *  :
nucleolin 1     GGFGGRGRGGFGGRGGFRGGRGGG

C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSGKLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
                     :* ::  * : ::  * :
nucleolin 2          GRGGFGGRGGGRGGGRGG

C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSGKLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
                    *:  ::  * : ::
nucleolin 3         KGEGGFGGRGGGRG

C1    VVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSGKLKGDDLQAIKKELTQIKQKVDSLLENLEKMEKE
                       : :*: :* ***  ::    ::*** :
poly A nuclease       QSSFSSPSAKSVNHSSSTLQTDDISVDK
```

FIGURE 10. Proteins that share primary sequence similarity to the basic and zipper region of C protein's bZLM and protein regions that share sequence similarity to the basic region alone. Shown above are proteins identified through an NCBI blastp database search that appear to share homology to regions of C1 between residues 140 and 207. Amino acid identity in the figure is indicated with a colon. Conservative substitutions are denoted with an asterisk. We have considered all apolar amino acids as conservative substitutions. Because a primary contribution to the free energy of binding for the carboxy-terminal binding region may come from hydrogen bonding, amino acids with this capability are also considered conserved (i.e., Ser and Arg would be conservative substitutions). Although many additional proteins were isolated in the search, the ones selected for presentation also contain consensus RNA recognition motifs. The proteins shown include C1 (human, residues 141–207), hnRNP A1 (human, residues 182–212), human auto-antigen p542 (human, residues 43–114), Raly (mouse, residues 127–198), nucleolin (Chinese hamster cells, domains 1 [residues 640–663], 2 [residues 624–641], and 3 [residues 613–626]), and poly A nuclease (yeast, residues 725–752). Although not indicated, poly-A nuclease has a three heptad repeat that begins 59 residues C-terminal to the denoted sequence.

Previous equilibrium binding studies conducted on 1–184 and 1–194 fragments of hnRNP A1 revealed that the latter had a 100-fold higher affinity for RNA (Shamoo et al., 1994). Residues 185–212 of hnRNP A1 have 68% homology (39% identity) with residues 143–174 of C protein. Thus, the residues responsible for enhancing the binding affinity of A1 (185–194) share homology with regions in C protein that are primarily responsible for C protein's RNA binding activity. Identification of a homologous region in another hnRNP that dramatically influences RNA binding affinity is consistent with this region functioning as a major determinant of RNA binding affinity in C protein. The presence of homologous regions in other nucleic acid binding proteins suggests that the basic region of C protein contains elements of an evolutionarily conserved RNA binding motif.

Two proteins identified in an NCBI Blastp database search, using Val 140–Glu 187 as the search sequence, were *Merc* (Duhl et al., 1994) and p542 (Vaughan et al., 1995a, 1995b) (Fig. 10). *Merc*, in addition to having homology to this region, has an RRM that is 80% homologous with the RRM of C protein. The function of *Merc* has not been determined, but embryonic mice cells lacking a functional gene encoding *Merc* do not develop beyond the blastocyst stage (Duhl et al., 1994). This developmental aberration has also been observed with mice that are homozygous for C protein gene knockouts (M. Roshon & E. Ruley, pers. comm.). p542 was identified originally as an auto-antigen in individuals infected with the Epstein Barr virus (Vaughan et al., 1995b). Auto-antibodies for p542 have now been found in a number of auto-immune disorders (Vaughan et al., 1995a). The extensive sequence homology between p542 and *Merc* suggests that they are homologous proteins. The homology between the two proteins extends beyond that depicted in Figure 10. *Merc*/p542 and C protein both possess a single amino-terminal RRM and a basic region leading into a leucine zipper. Moreover, the spacing of these motifs is the same. The presence of an amino-terminal basic region in these three proteins that precedes a leucine zipper is similar to domain organization in the basic-leucine zipper (bZIP) class of DNA binding proteins that have been identified in both prokaryotes and eukaryotes. Specific mutations within the leucine zipper of bZIP proteins can alter the state of oligomerization and/or the protein's binding properties (Harbury et al., 1993). Likewise, we have demonstrated that deletion of the leucine zipper in C protein abrogates oligomerization and alters the

RNA binding properties of the protein. Therefore, we believe that the basic region of *Merc*/p542 and C protein preceding a leucine zipper are evolutionarily and possibly functionally related to the bZIP DNA binding domain. We refer to this new "bZIP-like" motif as the bZLM.

It is interesting that 23% of the residues in basic region of the bZLM (Val 140–Glu 179) are serines. An analysis of the sequence alignments in Figure 10 suggests that there is also a conservation of the position of these residues between A1 and C protein. It has been shown previously that C protein is phosphorylated primarily at Ser residues and that hyperphosphorylation of the protein inhibits RNA binding (Mayrand et al., 1993). It has also been shown that, during mitosis, C protein is distributed in the cytoplasm in a hyperphosphorylated state (Pinol-Roma & Dreyfuss, 1993). Serines 143, 149, 153, 158, 169, 170, and 172 of the bZLM lie within recognition sites for protein kinase C and A. Phosphorylation of Ser 192 in hnRNP A1 by protein kinase C has been shown to inhibit RNA strand-annealing properties of the protein (Idriss et al., 1994). The homologous Ser of C protein's bZLM is in position 153, directly preceding a Lys-Ser-Gly repeat (described below). Because the bZLM is the primary determinant of C protein binding, it is likely that modulation of its affinity occurs through phosphorylation of one or more of the Ser residues described above.

A closer inspection of the bZLM reveals the presence of three repeats of the amino acid residues Lys-Ser-Gly. One of the repeats is conserved in hnRNP A1 and consists of the sequence Arg-Ser-Gly. A very interesting conservation of residues is observed between nucleolin and the Lys-Ser-Gly repeat of C protein. Specifically, the three peptide fragments of nucleolin shown in Figure 10 all contain the sequence Arg-Gly-Gly at many of the same positions occupied by Lys-Ser-Gly in C protein. The conservation of the spacing between the Lys-Ser-Gly repeat and the Arg-Gly-Gly repeat suggests that the bZLM of C protein may represent a variation of the RGG RNA binding motif. We are currently using site-specific mutagenesis to investigate the role of the tripeptide repeat and other residues in the bZLM region in RNA binding.

The binding affinity of the Val 140–Gly 290 variant to poly[r(ϵA)] is sufficient to account for all of the free energy of binding observed for the intact C protein. This suggests that the RRM in the wild-type protein may function more as an auxiliary RBD. In addition, this fragment displays high affinity for a SELEX-identified winner sequence compared with a randomized version of this winner, suggesting that the determinants for selecting the winner sequence also lie in this region. The affinity observed for the Val 140–Gly 290 protein exceeds that of the intact protein, suggesting that negative allosteric influences are removed by deleting residues 1–139. Because residues 1–104 contain the consensus RRM, it is possible that this region is an allosteric modulator of the bZLM. It is also likely that the bZLM in C protein is spatially close to the canonical RRM and that, in the intact protein, this region, along with the RRM, forms a single binding domain. Support for this latter hypothesis is derived from comparative analysis of the binding properties of the Met 1–Phe 115 and Met 1–Gly 159 polypeptides. The Met 1–Phe 115 variant binds RNA stoichiometrically in the absence of salt with an occluded site size of ~8 nt. We could not detect significant binding for this polypeptide at 100 mM NaCl. On the other hand, the Met 1–Gly 159 polypeptide binds poly[r(ϵA)] stoichiometrically at 100 mM NaCl with an occluded site size of ~8 nt. These polypeptides both bind poly[r(ϵA)] with identical site sizes and the only difference between them is residues 116–159 in the Met 1–Gly 159 variant. This region includes most of the 20 amino acids (140–161) that we have defined as being the minimal requirement for high-affinity salt-resistant binding. The identical site sizes observed for both proteins indicates that their mode of interaction with RNA is probably the same. A logical interpretation of the difference in affinity is that the 18 amino acids of the bZLM are proximally positioned relative to the RRM, and these provide additional interactions (i.e., hydrogen bonding, electrostatic, or hydrophobic) that further stabilize the RNA–protein complex. This particular model suggests that the bZLM and the RRM form one functional binding domain.

It has been reported that a deletion variant of C1 protein (residues 88–290) fused to a hemagglutinin tag does not bind poly rU linked to sepharose beads (Gorlach et al., 1994). The discrepancy between these and our findings could result from the former study conducting binding experiments in 1 M KCl. We have not explored the binding of any deletion variants under these solution conditions because ionic interactions are likely to be an important component of C protein's interaction with RNA. Additionally, it has been reported that C protein's enhanced affinity for specific SELEX-identified oligonucleotides is dependent on the 10-residue sequence carboxy-terminal to the RRM (residues 95–104) (Gorlach et al., 1994). As described here, however, the Val 140–Gly 290 construct (possessing the bZLM) displays especially high affinity for the winner oligonucleotide. Although it is surprising that two separate regions of C protein display enhanced affinity for the same substrate, it may indicate that the bZLM interacts with the amino-terminal RRM.

The bZLM region that we have identified in C protein clearly represents a new class of conserved RNA binding motif that either functions independently or as part of the consensus RRM. Because it exists at four copies per tetramer in one of the most abundant proteins in the eukaryotic nucleus that has been shown to bind RNA in a highly cooperative manner (McAfee

et al., 1996a), it is likely that it represents a motif in C protein that is involved in nonspecific RNA binding. As indicated above, isolated RRMs containing the 90–100 conserved amino acids may not represent "complete" RNA binding entities, but may interact with auxiliary regions to form a complete RBD. Structural and functional characterization of the bZLM will provide valuable new insights into the molecular basis of RNA–protein interactions.

MATERIALS AND METHODS

Cloning, expression, and purification of deletion mutants of C1

Cloning and expression of the cDNA encoding C1 using the published sequence of the gene are described elsewhere (Swanson et al., 1987; Burd et al., 1989; McAfee et al., 1996b). Deletion polypeptides of C1 were produced using the vector pET-3b (C1) (contains the entire coding region of C1 cloned into pET-3b) as a template for PCR amplification of specific DNA segments (see the Results and Fig. 1 for a description of recombinant C1 variants). Primers for the PCR reaction included either *Nde* 1 or *Bam*H 1 sites at their 5′ ends to facilitate ligation into the *Nde* 1/*Bam*H 1 site of the expression vector pET-28a (Novagen). The recombinant plasmid was then used to transform *E. coli* strain Bl21 DE3 (pLys S) (Studier et al., 1990). The expression of recombinant protein in this strain was as described by Studier et al. (1990). Purification of the resulting His-tag proteins were performed according to the manufacturer's instructions (Novagen).

Fluorescence binding experiments

The binding buffer used in all of the fluorescence experiments reported here contained 10 mM Tris, pH 8.0, and 1 mM EDTA. Unless stated otherwise, the NaCl concentration in this buffer was 100 mM. Fluorescence experiments were conducted using a SLM Aminco Bowman AB-2 Luminescence spectrometer. Excitation and emission slits were set at 4 and 8 nm, respectively. Poly[r(ϵA)] was generated by treatment of poly(A) with chloroacetaldehyde as described by Steiner et al. (1973). Fluorescence binding isotherms were generated by excitation at 310 nm and measuring the change in the emission at 410 nm as a function of increasing protein concentration. The fluorescence was corrected for dilution due to the addition of protein sample, and background fluorescence. The fluorescence enhancement was calculated using Equation 1:

$$E = \frac{(F - F_i)}{F_i}, \tag{1}$$

where E is enhancement, F is the fluorescence observed at each titration point, and F_i is the fluorescence observed in the absence of protein.

HPLC-purified oligoribonucleotides were purchase from Midland Certified Reagent or Promega and used without further purification. The concentration of each oligo was calculated using the nearest neighbor approximation method of Borer (1975). For competition assays, a control titration of poly[r(ϵA)] with the V140–G290 variant (see Fig. 1 for a description of this construct) was performed in the absence of competitor. Another titration was then performed with poly[r(ϵA)] and a SELEX-identified winner sequence (Gorlach et al., 1994) or a randomized version of the winner. An expression for the apparent equilibrium constant (K_{app}) for the interaction of the protein with each oligonucleotide was derived considering the following equilibria:

$$P + \epsilon A \leftrightarrow P\epsilon A$$

and

$$P + comp \leftrightarrow [Pcomp],$$

where P, ϵA, and $comp$ are the steady-state concentrations of protein, poly[r(ϵA)], and competitor RNA, respectively. $P\epsilon A$ and $Pcomp$ are the equilibrium concentrations of protein–ϵA or protein–competitor RNA complexes. The mass action expression defining the K_{app} for the interaction of the protein with poly[r(ϵA)] ($K_{\epsilon A}$) or competitor (K_{comp}) are:

$$K_{\epsilon A} = \frac{[P\epsilon A]}{[P]_{free}\,[\epsilon A]_{free}} \tag{2}$$

and

$$K_{comp} = \frac{[Pcomp]}{[P]_{free}\,[comp]_{free}}. \tag{3}$$

In a competition titration, the free protein concentration for both mass action expressions is the same. As a result, both expressions can be equated through $[P]_{free}$. The solution for K_{comp} then becomes Equation 4:

$$K_{comp} = \frac{[P_{comp}] \times K\epsilon A \times [\epsilon A]_{free}}{[P\epsilon A][comp]_{free}}. \tag{4}$$

In this equation, $K_{\epsilon A}$ is derived from a fit of data from the control titration (no competitor present) using the McGhee–von Hippel cooperative binding model (McGhee & von Hippel, 1974). In Figures 5, 6, 7, 8, and 9, protein concentrations are expressed in molar polypeptides.

Velocity sedimentation, CD, and T_m determinations

C1 deletion mutants were analyzed on 5–20% sucrose gradients as described by Bott et al. (1982). Ethanol-precipitated fractions collected from each gradient were analyzed by SDS-PAGE. Carbonic anhydrase (CA; 29 kDa), BSA (66 kDa), and immunoglobulin G (IgG; 150 kDa) were used as standards. Sedimentation coefficients were calculated by the methods described by Martin and Ames (1961). A synthetic peptide consisting of the 28 residues found at the leucine zipper region of C protein (Leu 180–Glu 207) with the addition of the tripeptide Cys-Gly-Gly at the amino-terminal end was prepared by solid phase methodology at Biosynthesis Inc. (Louisville, Texas). The Cys-Gly-Gly residues were added for use in future studies. The peptide was purified to >95% purity as determined by both HPLC and mass spectroscopy. CD spectra were acquired on a Jasco J-720 spectrometer at room temperature. The synthetic peptide (Leu 180–Glu 207) was prepared in PBS, pH 7.3, at concentrations of 7.4 mM (total residues) and 3.7 mM (total residues). The reduced samples were air oxidized by dialysis overnight. The peptide was then

subjected to Tris-tricine SDS-PAGE (Schagger & von Jagow, 1987) in the presence and absence of reducing agent (1%, v/v, β-mercaptoethanol in the sample buffer). The oxidized sample migrated slightly above aprotinin (6 kDa), as expected for a dimer (7,092 Da), whereas the reduced sample migrated as a monomer slightly above insulin. Ellipticity (θ) was measured from 190 to 250 nm for each scan. Spectra were then normalized to calculate the molar ellipticity ($[\theta]$) in order to allow for comparison. TFE was added to the appropriate samples at a ratio of 1:1 (v/v). Thermal stability of the Leu 180–Glu 207 peptide was monitored from 5 to 89 °C. Temperature was increased in steps of 2 °C. Samples were equilibrated for 1 min at each temperature before measuring the ellipticity from 190 to 250 nm.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM48567 (W.M.L.) and by a Ford Foundation Postdoctoral Fellowship 880-0765 (J.G.M).

Received March 5, 1996; returned for revision April 15, 1996; revised manuscript received August 19, 1996

REFERENCES

Amrute SB, Abdul-Manan Z, Pandey V, Williams KR, Modak MJ. 1994. Purification and nucleic acid binding properties of a fragment of type C1/C2 heterogeneous nuclear ribonucleoprotein from thymic nuclear extracts. *Biochemistry 33*:8282–8291.

Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EV. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci USA 88*:10148–10152.

Barnett SF, Friedman DL, LeStourgeon WM. 1989. The C proteins of HeLa 40S nuclear ribonucleoprotein particles exist as anisotropic tetramers of (C1)3C2. *Mol Cell Biol 9*:492–498.

Barnett SF, Theiry TA, LeStourgeon WM. 1991. The core proteins A2 and B1 exist as (A2)3B1 tetramers in 40S nuclear ribonucleoprotein particles. *Mol Cell Biol 11*:864–871.

Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci 4*:2455–2468.

Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. 1995. Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci USA 92*:8259–8263.

Beyer AL, Osheim YN. 1990. Ultrastructural analysis of the Ribonucleoprotein substrate for pre-mRNA processing. In: Strauss PR, Wilson SH, eds. *The eukaryotic nucleus: Molecular biochemistry and macromolecular assemblies, vol 2*. Caldwell, New Jersey: The Telford Press. pp 431–451.

Borer PN. 1975. Measured and calculated extinction coefficients at 260 nm, 25 °C and neutral pH for single stranded RNA. In: Fasman GD, ed. *Handbook of biochemistry and molecular biology*. Cleveland, Ohio: CRC Press. pp 359–360.

Bott RR, Navia MA, Smith JL. 1982. Improving the quality of protein crystals through purification by isoelectric focusing. *J Biol Chem 257*:9883–9886.

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

Burd CG, Swanson MS, Gorlach M, Dreyfuss G. 1989. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: A diversity of RNA binding proteins is generated by small peptide inserts. *Proc Natl Acad Sci USA 86*:9788–9792.

Chen J, Rajendran S, Lee JC, Matthews KS. 1994. Construction of a dimeric repressor: Dissection of subunit interfaces in Lac repressor. *Biochemistry 33*:1234–1241.

Chen YH, Yang JT, Chau KH. 1974. Determination of the helix and B form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*:3350–3359.

Cohen C, Parry DAD. 1986. Alpha-helical coiled coils—A widespread motif in proteins. *Trends Biochem Sci 11*:245–248.

Cohen C, Parry DAD. 1990. Alpha-helical coiled coils and bundles: How to design an alpha-helical protein. *Proteins Struct Funct Genet 7*:1–15.

Cohen C, Parry DAD. 1994. Alpha-helical coiled coils: More facts and better predictions. *Science 263*:488–489.

Conway G, Wooley J, Bibring T, LeStourgeon WM. 1988. Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. *Mol Cell Biol 8*:2884–2895.

Crick FHC. 1953. The packing of alpha-helices: Simple coiled-coils. *Acta Crystallogr 6*:689–697.

Duhl DMJ, Stevens ME, Vrieling H, Saxon PJ, Miller MW, Epstein CJ, Barsh GS. 1994. Pleiotropic effects of the mouse lethal yellow (Ay) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. *Development 120*:1695–1708.

Goldberg AL, Dice JF. 1974. Intracellular protein degradation in mammalian and bacterial cells. *Annu Rev Biochem 43*:835–869.

Goldberg AL, St. John AC. 1976. Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annu Rev Biochem 45*:747–803.

Gorlach M, Burd CG, Dreyfuss G. 1994. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem 269*:23074–23078.

Gottesman S, Maurizi MR. 1992. Regulation by proteolysis: Energy-dependent proteases and their targets. *Microbiol Rev 56*:592–621

Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science 262*:1401–1407.

Herschlag D. 1995. RNA chaperones and the RNA folding problem. *J Biol Chem 270*:20871–20874.

Ho YS, Mahoney ME, Wulff DL, Rosenberg M. 1988. Identification of the DNA binding domain of the phage λ *c*II transcriptional activator and the direct correlation of *c*II protein stability with its oligomeric forms. *Genes & Dev 2*:184–195.

Hoffman DW, Query CC, Golden BL, White SW, Keene JD. 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA 88*:2495–2499.

Huang M, Rech JE, Northington SJ, Flicker PF, Mayeda A, Krainer AR, LeStourgeon WM. 1994. The C protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol Cell Biol 14*:518–533.

Idriss H, Kumar A, Casas-Finet JR, Guo H, Damuni Z, Wilson SH. 1994. Regulation of in vitro nucleic acid strand annealing activity of heterogeneous nuclear ribonucleoprotein protein A1 by reversible phosphorylation. *Biochemistry 33*(37):11382–11390.

Keene JD, Query CC. 1991. Nuclear RNA-binding proteins. In: Moldave K, Cohn W, eds. *Progress in nucleic acid research and molecular biology, vol. 41*. New York: Academic Press. pp 179–202.

Kiledjian M, Burd CG, Gorlach M, Portman DS, Dreyfuss G. 1994. Structure and function of hnRNP proteins in RNA protein interactions. In: Mattaj I, Nagai K, eds. *Frontiers in molecular biology*. Oxford, UK: Oxford University Press. pp 127–149.

Kouzarides T, Ziff E. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature 336*:646–651.

Landschulz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science 240*:1759–1763.

Lee AL, Kanaar R, Rio DC, Wemmer DE. 1994. Resonance assignments and solution structure of the second RNA-binding domain of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry 33*:13775–23786.

LeStourgeon WM, Barnett SF, Northington SJ. 1990. Tetramers of the core proteins of 40S nuclear ribonucleoprotein particles assemble to package nascent transcripts into a repeating array of regular particles. In: Strauss PR, Wilson SH, eds. *The eukaryotic nucleus: Molecular biochemistry and macromolecular assemblies, vol 2*. Caldwell, New Jersey: The Telford Press. pp 477–502.

LeStourgeon WM, Lothstein L, Walker BW, Beyer AL. 1981. The com-

position and general topology of RNA and protein in monomer 40S ribonucleoprotein particles. In: Busch H, ed. *The cell nucleus, vol 9*. New York: Academic Press. pp 49–87.

Lu J, Hall KB. 1995. An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J Mol Biol 247*:739–752.

Lumb KJ, Carr CM, Kim PS. 1994. Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4. *Biochemistry 33*:7361–7367.

Martin RG, Ames BN. 1961. A method for determining the sedimentation behavior of enzymes: Application to protein mixtures. *J Biol Chem 236*:1372–1379.

Mayrand SH, Dwen P, Pederson T. 1993. Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. *Proc Natl Acad Sci USA 90*:7764–7768.

McAfee JG, Huang M, Soltaninassab S, Rech JE, Iyengar S, LeStourgeon WM. 1996a. The packaging of pre-mRNA. In: Krainer A, ed. *RNA processing, frontiers in molecular biology*. Oxford, UK: Oxford University Press. Forthcoming.

McAfee JG, Soltaninassab SR, Lindsay ME, LeStourgeon WM. 1996b. Proteins C1 and C2 of heterogeneous nuclear ribonucleoprotein complexes bind RNA in a highly cooperative fashion: Support for their contiguous deposition on pre-mRNA during transcription. *Biochemistry 35*:1212–1222.

McGhee JD, von Hippel PH. 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

Merrill BM, Barnett SF, LeStourgeon WM, Williams KR. 1989. Primary structure differences between proteins C1 and C2 of HeLa 40S nuclear ribonucleoprotein particles. *Nucleic Acids Res 17*:8441–8449.

Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature 348*:515–520.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science 254*:539–544.

O'Shea EK, Rutkowski R, Kim PS. 1989a. Evidence that the leucine zipper is a coiled coil. *Science 243*:538–542.

O'Shea EK, Rutkowski R, Stafford WF, Kim PS. 1989b. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science 245*:646–648.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 angstrom resolution of the RNA-binding domain of the U1A spliceosomal protein complexes with an RNA hairpin. *Nature 372*:342–438.

Pinol-Roma S, Dreyfuss G. 1993. Cell cycle-regulated phosphorylation of the pre-mRNA-binding (heterogeneous nuclear ribonucleoprotein) C proteins. *Mol Cell Biol 9*:5762–5770.

Platt T, Miller JH, Weber K. 1970. In vivo degradation of mutant Lac repressor. *Nature 228*:1154–1156.

Rech JE, Huang MH, LeStourgeon WM, Flicker PF. 1995a. An ultrastructural characterization of in vitro-assembled hnRNP C protein–RNA complexes. *J Struct Biol 114*:84–92.

Rech JE, LeStourgeon WM, Flicker PF. 1995b. Ultrastructural morphology of the hnRNP C protein tetramer. *J Struct Biol 114*:77–83.

Ross A, Laws JB, Rousslang WR, Wyssbrod HR. 1992. Tyrosine fluorescence and phosphorescence from proteins and polypeptides. In: Lakowicz JR, ed. *Topics in fluorescence spectroscopy*. New York: Plenum Press. pp 1–53.

Saudek V, Pasley HS, Gibson T, Gausepohl H, Frank R, Pastore A. 1991. Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry 30*:1310–1317.

Schagger H, von Jagow G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDA. *Anal Biochem 166*:368–379.

Schuermann M, Neuberg M, Hunter JB, Jenuwein T, Ryseck RP, Bravo R, Muller R. 1989. The leucine repeat motif in Fos protein mediates complex formation with Jun/AP-1 and is required for transformation. *Cell 56*:507–516.

Shamoo Y, Abdul-Manan N, Patten AM, Crawford JK, Pellegrinik MC, Williams KR. 1994. Both RNA-binding domains in heterogeneous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry 33*:8272–8281.

Steiner RF, Kinner W, Lunasin A, Delac J. 1973. Fluorescent derivatives of polyribonucleotides containing e-adenosine. *Biochem Biophys Acta 294*:24–37.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Swanson MS, Nakagawa TY, LeVan K, Dreyfuss G. 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: Conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-RNA binding proteins. *Mol Cell Biol 7*:1731–1739.

Vaughan JH, Nguyen MD, Valbracht JR, Patrick K, Rhodes GH. 1995a. Epstein-Barr virus-induced autoimmune response: II. Immunoglobulin G autoantibodies to mimicking and non mimicking epitopes. Presence in autoimmune disease. *J Clin Invest 95*:1316–1327.

Vaughan JH, Valbracht JR, Nguyen MDH, Smith RS, Patrick K, Rhodes GH. 1995b. Epstein-Barr virus-induced autoimmune responses. *J Clin Invest 95*:1306–1315.

Vinson CR, Sigler PB, McKnight SL. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science 246*:911–916.

Wittekind M, Gorlach M, Friedrichs M, Dreyfuss G, Mueller L. 1992. 1H, 13C, and 15N NMR assignments and global folding patterns of the RNA-binding domain of the human hnRNP C protein tetramer. *Biochemistry 31*:6254–6265.

Zhou NE, Kay CM, Hodges RS. 1992. Synthetic model proteins: The relative contribution of leucine residues at the nonequivalent positions of the 3–4 hydrophobic repeat to the stability of the two-stranded α-helical coiled-coil. *Biochemistry 31*:5739–5746.