Purification and Nucleic Acid Binding Properties of a Fragment of Type C1/C2 Heterogenous Nuclear Ribonucleoprotein from Thymic Nuclear Extracts†

Sheela B. Amrute,‡ Zehan Abdul-Manan,§ Virendranath Pandey,‡ Kenneth R. Williams,∥ and Mukund J. Modak*,‡

Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School, Newark, New Jersey 07103, and Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06510

Received January 27, 1994; Revised Manuscript Received April 29, 1994®

ABSTRACT: A single-strand nucleic acid binding protein (C/F) that has an apparent molecular weight of 12 000 on SDS–polyacrylamide gel electrophoresis and that was originally thought to be the 12-kDa α-subunit of the AB form of terminal deoxynucleotidyl transferase (TdT) from calf thymus has been purified and identified as a fragment of the type C1/C2 hnRNP proteins. On the basis of NH_2-terminal sequencing and mass spectrometric analysis, C/F contains ~94 residues and spans from residue 9 to approximately residue 102 in the type C1/C2 hnRNP proteins. C/F is presumably produced *in vitro* via limited proteolysis of the type C1/C2 hnRNP proteins following cell disruption. Since C/F corresponds almost exactly to the ~90-residue conserved ribonucleoprotein binding domain (RBD) that is shared by many eukaryotic RNA binding proteins, it provided an opportunity to better characterize the domain structure of the type C1/C2 hnRNP proteins and to compare the nucleic acid binding properties of the type C1/C2 and A1 [see Shamoo *et al.* (1994) *Biochemistry*, preceding paper in this issue] RNA binding domains. Like the type A1 RBD, the type C1/C2 RBD has an apparent occluded site size of 6–7 nucleotides. The type C1/C2 RBD binds non-cooperatively to homopolynucleotides and has preferential affinity for RNA and for single as opposed to double-stranded nucleic acids. The type C1/C2 RBD has about a 100-fold higher affinity than the type A1 RBD does for RNA and some of this increased affinity results from additional ionic interactions. The latter account for ~50% of the free energy of binding of the type C1/C2 RBD. While the type C1/C2 hnRNP proteins exist *in vivo* as a very tight tetramer with the structure $(C1)_3C2$ [Barnett *et al.* (1989) *Mol. Cell. Biol. 9*, 492–498], the isolated type C1/C2 RBD is a monomer. Hence, the determinants for tetramerization appear to lie outside the type C1/C2 RBD. Phenylalanine 19 was identified as the only point of photochemical cross-linking of the type C1/C2 RBD to $[d(T)]_8$. This residue corresponds to the major site of cross-linking of the A1 RBD to $[d(T)]_8$ [Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., & Williams, K. R. (1988) *J. Biol. Chem. 263*, 3307–3313]. Thus, even though the overall extent of direct sequence identity between the type A1 and C RNA binding domains is only 17%, both proteins appear to share a topologically similar RNP:oligonucleotide interface which does not appear to be substantially altered when it is placed within the context of a multidomain hnRNP protein.

In the preceding paper in this issue we characterized the nucleic acid binding properties of fragments of the A1 hnRNP[1] corresponding to its two RNA binding domains (residues 1–92 and 93–184), the region spanning both these domains (residues 1–184), and an overlapping fragment that has been referred to as the UP1 protein (residues 1–196). This approach succeeded in determining the relative contributions of these individual domains to A1 binding to a fluorescent, single-stranded RNA. In this paper, these binding studies are extended to include the corresponding RNA binding domain (RBD) in the type C1/C2 hnRNP.

Like the type A and B hnRNP proteins, the type C1/C2 proteins are abundant hnRNP proteins that contain the ~90-residue, conserved eukaryotic RNA binding domain (RBD) that has now been found in numerous other proteins such as the poly(A) binding protein, nucleolin, and the *Drosophila* differentiation proteins sxl and tra-2 that share the common property of binding RNA (Bandziulis *et al.*, 1989; Merrill & Williams, 1990; Kenan *et al.*, 1991). The domain structure of the type A and B hnRNP proteins is similar in that each contains two RBDs that are arranged in tandem at their NH_2-terminus and that is followed by a glycine-rich, carboxy-terminal domain. The type C1/C2 hnRNP contain a single RBD at their NH_2-terminus that is followed by an acidic, COOH-terminal domain. At least some of the diversity of the hnRNP family of proteins appears to result from alternative splicing. Hence, B1 differs from A2 only by the presence of a 12-amino acid insert near its amino terminus (Burd *et al.*, 1989). Similarly, C2 differs from C1 by the presence of a 13-amino acid insert that is located just after the RBD region in C1 (Burd *et al.*, 1989; Merrill *et al.*, 1989). Since the type C1/C2 proteins exist as a very tight, anisotropic tetramer that has the structure $(C1)_3C2$ (Barnett *et al.*, 1989), this

† Supported in part by grants from NSF (M.J.M. DMB-87-15829) and NIGMS (K.R.W. GM 31539 and M.J.M. GM 36307) and by a fellowship from the New Jersey Commission on Cancer Research (S.B.A.).

‡ Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School.

§ Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine.

∥ Howard Hughes Medical Institute, Yale University School of Medicine.

® Abstract published in *Advance ACS Abstracts,* June 15, 1994.

[1] Abbreviations: TdT, terminal deoxynucleotidyl transferase; C/F, type C1/C2 protein fragment; poly(U), polynucleotide uracil; oligo(dT), deoxyoligonucleotide thymidine; hnRNP, heterogenous ribonucleoprotein particles, kDa, kilodalton; DEAE, diethylaminoethyl; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; PC, phosphocellulose; EDTA, ethylenediaminetetraacetic acid; PABP, poly(A) binding protein; TFA, trifluoroacetic acid; TCA, trichloroacetic acid; UV, ultraviolet; ssDNA, single-stranded DNA; HPLC, high-performance liquid chromatography; UP1, unwinding protein 1.

0006-2960/94/0433-8282$04.50/0

13-amino acid insert in C2 may play an important role in protein:protein interactions within this complex (Merrill *et al.*, 1989). Although hnRNP particle reconstitution studies fail to detect any base- or sequence-specific binding by hnRNP proteins (Wilk *et al.*, 1983; LeStourgeon *et al.*, 1990), an RNase protection study suggests specific, high-affinity A1 and type C1/C2 protein binding to the 3′-end of introns within a region containing the conserved poly(pyrimidine) stretch between the branch site and the 3′-splice site (Swanson & Dreyfuss, 1988). Similarly, the type C1/C2 proteins have been shown to have unusually high affinity for poly(U) sequences (Swanson & Dreyfuss, 1988; Wilusz *et al.*, 1988).

In this report, we describe the purification of an ~10 kDa protein (which migrates with an apparent molecular weight of 12 kDa on SDS–PAGE) from extracts of calf thymus chromatin that, because of its similar properties, was mistaken for the 12-kDa α-subunit of the AB form of terminal deoxynucleotidyl transferase (TdT) (Pandey & Modak, 1987). Although both proteins copurify through several steps and bind tightly to single-stranded oligomeric DNA, separation can be achieved on DEAE-Sephadex. Since subsequent amino acid sequencing demonstrated that this 10-kDa protein is actually a fragment corresponding to the type C1/C2 RBD domain, we have called this protein C/F. Since the C/F protein afforded an ideal opportunity to compare the nucleic acid binding properties of the isolated type C1/C2 RBD with that reported in the literature for the tetrameric type C1/C2 protein complex, we have characterized the nucleic acid binding properties of C/F by utilizing photochemical cross-linking as well as fluorescence assays. Now that the nucleic acid binding properties of the A1 hnRNP and its fragments have been characterized (see preceding paper in this issue), comparisons between the binding properties of the A1 and C1/C2 type proteins can be made. Results from these studies correlate well with the apparent higher affinity of the C1/C2 type as compared to the A-type hnRNP for 40S hnRNP particles (Beyer *et al.*, 1977). Furthermore, studies on the A1 and C1/C2 type proteins provide good prototypical models for understanding the role of single versus multiple RBDs in binding and contribute to our general understanding of the molecular basis for the interaction of RBD domains with nucleic acids.

MATERIALS AND METHODS

[^{3}H]- and [^{32}P]-labeled deoxynucleoside triphosphates (dNTPs), ribonucleoside triphosphates (rNTPs), and synthetic oligo- and polydeoxynucleotides were obtained from P. L. Biochemicals. Oligo(U) was prepared by brief exposure of poly(U) to 0.1 M NaOH followed by neutralization. The average length of the oligo(U) was between 10 and 20 nucleotides. Calf thymus DNA (Sigma Chemical Co.), activated with pancreatic DNase I to a level of 20% acid solubility as described by Edenbert *et al.* (1978), was used as a primer for TdT assays. SDS–polyacrylamide gel electrophoresis reagents were purchased from Sigma Chemical Co. and electrophoresis was performed using the Laemmli system (Laemmli, 1970). The various column chromatography resins were commercial products obtained as follows: phosphocellulose P-11 from Whatman, DEAE-Sephadex A-50 from Pharmacia, and oligo[d(T)]-cellulose (20003) from Collaborative Research, Inc. T4 polynucleotide kinase was purchased from Bethesda Research Lab.

Enzyme Assays. TdT was assayed as described before using both activated DNA and synthetic template primer (Modak, 1978, 1979). Detection of endo- and exonuclease activity was carried out as follows: A commercial preparation of $[d(A)]_{20}$ was labeled at its 5′-end by polynucleotide kinase, essentially as described by Bacerra and Wilson (1984). Reaction mixtures containing 50 mM Tris-HCl, pH 7.6, 3 nmol of [5′-^{32}P]$[d(A)]_{20}$ (specific radioactivity 50 cpm/pmol), 20 μg of polynucleotide kinase, and 1 mM Mn^{2+} or 2.5 mM Mg^{2+} were incubated at 37 °C for 1 h. At the end of the incubation, an aliquot of the reaction was analyzed on a 20% polyacrylamide gel containing 8 M urea (Bacerra & Wilson, 1984).

Assay for C/F. C/F was detected by its ability to covalently cross-link with [5′-^{32}P]$[d(T)]_8$ upon brief exposure to UV irradiation (Modak & Gillerman-Cox, 1982; see Williams & Konigsberg, 1992, for a review of photo-cross-linking). Typically, a 50-μL reaction mixture contained 26 pmol [5′-^{32}P]$[d(T)]_8$ (45 000 Cerenkov cpm), 50 mM Tris-HCl, pH 7.8, and a 5–20 μL aliquot of the protein fraction to be assayed. The reaction mixture was exposed to UV irradiation (254 nm, 200 μW/cm^2 output at 5 cm) in polypropylene tubes at a distance of 4 cm for 12 min. The covalently cross-linked product was subjected to 12% polyacrylamide–SDS gel electrophoresis (Laemmli, 1970), followed by autoradiography. The TCA precipitation assay was carried out as follows: UV irradiated, cross-linked protein–$[d(T)]_8$ complexes were precipitated with 10% TCA, collected on glass fiber filters (GF/B), washed, and then counted for radioactivity.

Purification of C/F. Since C/F copurifies with TdT, the initial procedures used in the extraction and purification of this protein are the same as that for TdT (Pandey & Modak, 1987). Frozen calf thymus gland (500 g) was cut into small pieces and homogenized in 3 volumes of buffer A (10 mM Tris-HCl, pH 7.8, 0.1 mM EDTA, 1 mM β-mercaptoethanol, 1 mM phenylmethanesulfonyl fluoride, 10% glycerol) supplemented with 0.5% Nonidet P40. The homogenate was filtered through cheesecloth and centrifuged at 5000*g* for 30 min. The cytoplasmic extract did not contain detectable $[d(T)]_8$ binding activity in the mass range of 10–15 kDa. All such activity was instead associated with the chromatin pellet and could be quantitatively extracted with 0.4 M NaCl. The cytoplasmic fraction was therefore discarded and the chromatin pellet was extracted with 3 volumes of buffer A containing 0.5 M NaCl. Nonidet P40 was omitted from all future buffers. The chromatin extract was centrifuged at 12000*g* for 60 min and the clear supernatant (diluted 6-fold with buffer A) was mixed with 400 mL of phosphocellulose pre-equilibrated with buffer A. The phosphocellulose was then packed in a column and washed with 10 volumes of buffer A containing 0.1 M NaCl, followed by 3 volumes of 0.2 M NaCl and finally by 3 volumes of 0.5 M NaCl. All activity was contained within the 0.5 M eluate. Active fractions containing C/F (as well as TdT) were concentrated by stepwise ammonium sulfate addition. The precipitate obtained at a concentration of 35% salt was discarded and the one obtained at 80% saturation was dissolved in buffer A and dialyzed against the same buffer. The dialysate was clarified by centrifugation and applied to a DEAE-Sephadex A-50 column pre-equilibrated with buffer A. At this stage, TdT is retained on the column whereas C/F is eluted in the flow-through fractions. The combined DEAE-Sephadex flow through and buffer wash fractions were heated at 60 °C for 3 min followed by chilling of the flask in an ice water bath. The heat-denatured protein precipitate was removed by centrifugation and the supernatant containing active protein was adsorbed onto a 1-mL column of oligo[d(T)]-cellulose, pre-equilibrated with 0.1 M NaCl in buffer A, at a flow rate of 7 mL/h. The

column was washed with 25 mL of 0.1 M NaCl in buffer A, followed by 8 mL each of buffer A containing 0.2, 0.5, 1 and 2 M NaCl and 2-mL fractions were collected. The column was then transferred to a 50 °C water bath and washed with 10 mL of buffer A preheated to 50 °C, followed by 25 mL of buffer A containing 0.5 M salt at the same temperature. Washing of the column and elution of the protein at the elevated temperature was carried out at a flow rate of 20 mL/h and 2-mL fractions were collected.

Active fractions were pooled, concentrated via a Centricon 10 microconcentrator, adjusted to 0.1% TFA, and injected onto a Vydac C-4 reverse-phase column (0.45 × 25 cm, 5-μm particle size and 300-Å pore size) that had been pre-equilibrated with 0.1% TFA (solvent A). Elution of proteins, at a flow rate of 0.7 mL/min, was effected by increasing the concentration of solvent B (70% acetonitrile in 0.1% TFA) according to the following time program: 0–25% B (0–20 min), 25–70% B (20–40 min), 70–100% B (40–60 min). HPLC analyses were carried out on a Varian Vista 5500 HPLC system equipped with a polychrome 9060 diode-array detector.

Amino Acid Composition. Amino acid analyses were carried out on a Beckman Model 6300 analyzer following a 16-h hydrolysis at 115 °C in 6 N HCl, 0.2% phenol. Under the analysis conditions used, proline and cysteine coelute and it is not possible to accurately quantitate tryptophan, as it is partially destroyed by the hydrolysis.

Cyanogen Bromide Cleavage, HPLC Peptide Separation, and Amino Acid Sequencing. Peak 2 (C/F, 7.1 μg ca. 750 pmol) was cleaved with cyanogen bromide by incubating with a 2000-fold molar excess of this reagent in 70% formic acid for 24 h at room temperature. Following cleavage, 33% of the reaction mixture was subjected to amino terminal sequencing. The remaining solution was dried *in vacuo* prior to dissolving in 200 μL of 6 M guanidine hydrochloride and injecting onto a Vydac C-4, 300-Å, 5-μm particle size column that measured 0.46 × 25 cm and that had been equilibrated with 0.05% trifluoroacetic, 1.6% acetonitrile. Peptides were eluted with increasing concentrations of buffer B (ca. 0.05% trifluoroacetic acid, 80% acetonitrile) as follows: 0–63 min (2–37% B), 63–95 min (37–75% B), 95–105 min (75–98% B). Those peptides selected for sequencing were directly loaded onto an Applied Biosystems Model 470 or 477 protein/peptide sequencer that was equipped with an "on-line" HPLC and that was operated according to the manufacturer's instructions.

Preparative-Scale Cross-Linking of C/F to [^{32}P][d(T)]$_8$, Tryptic Digestion, and Isolation of the [d(T)]$_8$ Cross-Linked Peptide. Six nanomoles (~75 μg) of C/F was mixed with 18 nmol of [^{32}P][d(T)]$_8$ in a final volume of 500 μL containing 50 mM Tris-HCl, pH 7.8, 0.1 mM EDTA, 5 mM mercaptoethanol, and 10% glycerol. The mixture was then distributed in 50-μL aliquots into parafilm wells and was exposed to UV irradiation as described in the assay procedure. Irradiated samples were pooled, chilled on ice for 30 min, and mixed with 50 μL of 100% TCA solution. The precipitated protein was recovered by centrifugation and was then washed twice with each of 10% TCA, 95% acetone, and 100% acetone. It was then dissolved in 0.4 M ammonium bicarbonate containing 8 M urea followed by incubation at 50 °C for 15 min. After diluting with water to 2 M urea, trypsin (TPCK-treated) was added at a ratio of 1:25 (trypsin:C/F, w:w) and digestion was carried out overnight at 37 °C. The digest was diluted to 0.5 M urea and was then loaded onto a 600-μL column of DEAE-Sephadex A-50, which was equilibrated with 20 mM TEAA buffer (pH 7.0). The column was then washed with increasing concentrations of TEAA to effect differential elution of the peptides. The radioactive (cross-linked) peptides eluted at 0.5 M TEAA. These fractions were collected, lyophilized, and desalted on a C-4 matrix. The matrix bound peptide was eluted with 75% methanol in 5 mM TEAA, lyophilized to dryness, and subjected to amino acid sequencing. The recovery of peptide from the C-4 matrix, as judged by the radioactivity associated with it, was determined to be ~50%.

Stoichiometry of Binding of C/F to ssDNA. Measurement of the binding stoichiometry of C/F to ssDNA was carried out by estimating protection of [^{3}H]-labeled M13mp ssDNA (O'Donnell *et al.*, 1987) from DNAse I. *Escherichia coli* K_{12} strain JM101 infected with M13mp18 bacteriophage was grown in medium containing 50 μg/mL of [^{3}H]thymidine (18.5 Ci/mol) as described by Miller (1972). The [^{3}H]-labeled M13mp18 ssDNA was isolated as described by Schreier and Cortese (1979). The purified [^{3}H]-labeled phage DNA was diluted with unlabeled phage DNA to a specific activity of ca. 2 × 10^5 cpm/μg of DNA. The reaction mixture (100 μL) containing 50 mM Tris-HCl, pH 7.4, 0.1 mM DTT, 1.0 μg of [^{3}H]-labeled ssDNA (2 × 10^5 cpm), and variable concentrations of C/F was incubated for 10 min at 37 °C, after which was added 5 μL of a solution containing 0.1 mg/mL each of DNase I and snake venom phosphodiesterase. The reaction mixtures were incubated for another 5 min prior to supplementing with 50 μL of 100 mM pyrophosphate solution containing 5 μg of sonicated calf thymus DNA. The amount of DNA protected by C/F from nuclease digestion was determined by TCA precipitation. A reaction mix containing all the above components except C/F was used as a control. In the absence of C/F, less than 0.1% of the labeled DNA is acid precipitated and retained on the filter. The saturation point was estimated by plotting the amount of acid precipitable radioactivity versus the C/F concentration. Since this assay neither takes into account non contiguous binding by C/F (thus leading to an overestimate of the actual concentration of bound lattice and, hence, the site size) nor does it permit an independent measure of the free versus bound C/F (thus leading to an overestimate of the concentration of bound C/F and, hence, an underestimate of the site size), it provides only an approximate estimate of the occluded site size.

Fluorescence Spectroscopy. Fluorescence titrations were carried out and the resulting data interpreted as described in the preceding paper (Shamoo *et al.*, 1994). For poly[r(εA)] titrations that were carried out in the presence of nucleic acid competitors, it was assumed that the free protein concentration is identical at the same extent of fluorescence enhancement of poly[r(εA)] in the absence or presence of a nucleic acid competitor. This assumption allows for the determination of the apparent affinity of the competing polynucleotide (K_{comp}) using the expression

$$K_{comp} = [comp]_{bound}[poly[r(\epsilon A)]]_{free}K_{poly[r(\epsilon A)]}/\{[comp]_{free}[poly[r(\epsilon A)]]_{bound}\}$$

In this expression, the concentration of the bound competitor is calculated from the difference between the total C/F concentration and the sum of the free and poly[r(εA)]-bound C/F (at that point in the titration) with the free C/F concentration being derived from an identical titration carried out in the absence of competitor.

RESULTS

Since C/F was originally detected as a contaminant of terminal deoxynucleotidyl transferase, the early steps for its

Table 1: Purification of C/F from Chromatin Extract of Calf Thymus Glands

step	procedure	total protein (mg)	specific activity[a] (units/mg)
1	0.5 M NaCl chromatin extract from 400 g of thymus gland[b]	4050	–
2	phosphocellulose chromatography	580	10
3	DEAE-Sephadex A-50 chromatography (flow through and buffer was following heat treatment)	25	145
4	oligo[d(T)]-cellulose	0.4	6165
5	C-4 HPLC	0.3	6579

[a] One unit of C/F is defined as the amount of protein required to bind 1 pmol of $[d(T)]_8$ (as judged by UV-mediated cross-linking). [b] Chromatin extract was found to contain a high level of endo- and exonuclease activity, and hence, the specific activity of C/F in the extract could not be estimated.

extraction from chromatin and column chromatographic purification were those used in TdT purification (Pandey & Modak, 1987). We determined that cytoplasmic extracts did not contain detectable C/F activity. Furthermore, C/F appears to be quantitatively extracted from chromatin with 0.5 M salt. The addition of detergent to the extraction buffer is unnecessary as it did not increase the yield of C/F. As described in Table 1, the C/F purification involves adsorption of chromatin extracts onto phosphocellulose followed by DEAE-Sephadex chromatography of the active fractions. The heat treatment of the DEAE fractions together with the above two column steps achieves nearly 160-fold purification.

Oligo[d(T)]-Cellulose and C-4 Reverse-Phase Chromatography. ssDNA cellulose matrix could not be used for further purification since C/F activity eluted in all fractions between 0.2 M and 0.8 M salt. Since C/F has consistently shown high affinity for oligo[*d(T)*], we attempted to employ oligo[*d(T)*]-cellulose for further purification. However, when this column was eluted at 4 °C, the recovery of C/F activity was extremely poor. That is, less than 20% of the activity was released in all salt washes (including 2 M) combined and no additional activity was recovered by altering the pH of the eluting buffer. However, increasing the column temperature to 50 °C resulted in 68% of the C/F activity eluting in the 0.5 M NaCl wash. This purified C/F did not have any detectable exo or endonucleolytic activity (data not shown). When the oligo[d(T)]-cellulose purified C/F was subjected to SDS–PAGE or Sephadex gel filtration, a single stainable protein and activity band corresponding to a molecular weight marker of ~12 kDa (cytochrome C) was obtained (data not shown). However, this fraction was resolved into two peaks upon C-4 reverse-phase HPLC. As shown in Figure 1, these peaks eluted at 39 min (68% B) and 44 min (77% B) in about a 1:2 ratio, respectively. Since both peaks have similar amino acid compositions (data not shown), we suspect they represent similar C/F species differing only with respect to cleavage site, extent of posttranslational modification, and/or other subtle changes such as one or more amino acid substitutions. Except where noted, we have used the protein contained in peak 2 from the C-4 column.

Protein Chemistry Studies on C/F. While both absorbance peaks obtained from the C-4 column have similar amino acid compositions and apparent affinities for $[d(T)]_8$ (see below), they can be readily differentiated on the basis of their masses as determined by laser desorption mass spectrometry (LDMS). Hence, whereas LDMS analysis of peak 1 gave two peaks of approximately equal intensity corresponding to masses of 10 342 and 10 802, respectively, only a single major peak of mass 10 133.7 was detected for peak 2. The identity of C/F

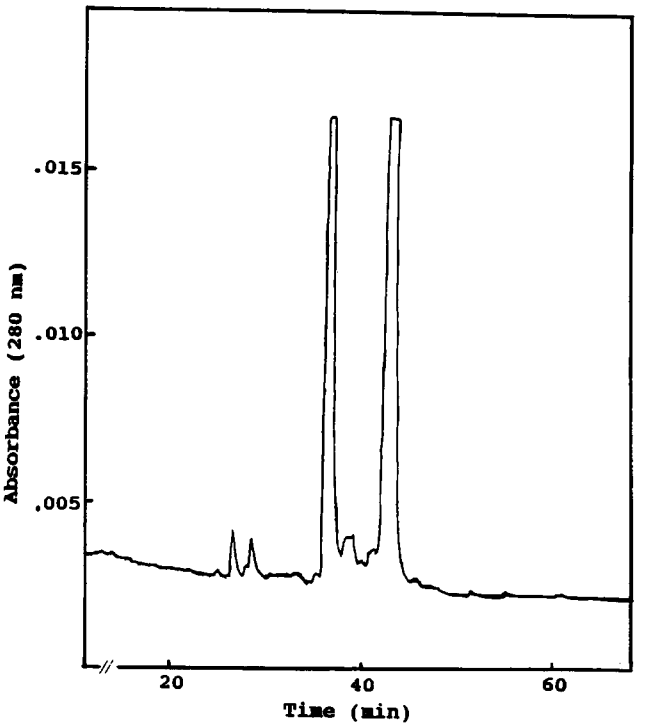

FIGURE 1: Reverse-phase HPLC profile of C/F purified through $[d(T)]_8$-cellulose. C/F, prepared as described in Materials and Methods, was injected onto a Vydac C-4 reverse-phase column equilibrated with 0.1% TFA and eluted with increasing concentrations of acetonitrile as described in Materials and Methods. Fractions were collected every minute and assayed for C/F activity by the TCA precipitation method described in Materials and Methods.

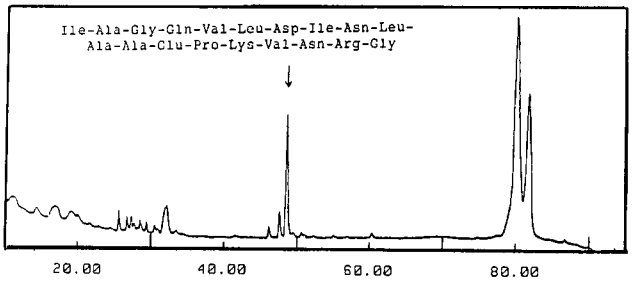

FIGURE 2: HPLC analysis of CNBr peptides derived from C/F protein and the amino acid sequence of one of these peptides. The reverse-phase HPLC separation was carried out on 760 pmol of cyanogen bromide-digested C/F protein. After cleavage, the reaction mixture was dried *in vacuo* prior to dissolving in 200 μL of 6 M guanidine hydrochloride and injecting onto a Vydac C-4 column (4.6 × 250 mm) that was eluted as described in Materials and Methods. Peptides were detected by their absorbance at 210 nm and the *x*-axis above is in terms of minutes. As described in Materials and Methods, the peak eluting at about 49 min was subjected to amino acid sequencing which resulted in assigning the 19 residues given above.

was established by amino acid sequencing of one of its cyanogen bromide fragments. As shown in Figure 2, reverse-phase HPLC of a cyanogen bromide digest of C/F resolved three major peaks. Amino acid sequencing of the peak eluting at 49 min provided 19 residues of sequence that exactly matched residues 75–94 of the type C1/C2 hnRNP protein (Burd *et al.*, 1989). An additional C/F sequence was obtained by amino acid sequencing of another sample of cyanogen bromide-digested protein that had not been subjected to HPLC fractionation. Three sequences were obtained that could be differentiated by matching to the known type C1/C2 sequence. These sequences began at threonine 9, asparagine 15, and isoleucine 75. Since threonine 9 in the type C1/C2 sequence does not follow a methionine, this must represent the NH_2-

Table 2: Amino Acid Sequencing of $[d(T)]_8$: C/F Cross-Linked Peptides

cycle number	amino acid identified[a] major peak	amino acid identified[a] minor peak	corresponding residue number in type C1/C2 protein major peak	corresponding residue number in type C1/C2 protein minor peak
1	Ser	Val	13	18
2	Met	*b*	14	19
3	Asn	Ile	15	20
4	Ser	Gly	16	21
5	Arg	Asn	17	22
6	Val	Leu	18	23
7	*b*	Asn	19	24
8	Ile	Thr	20	25
9	Gly	Leu	21	26
10	Asn	Val	22	27
11	Leu	Val	23	28
12	Asn	Lys	24	29
13	Thr		25	
14	Leu		26	
15	Val		27	
16	Val		28	
17	Lys		29	

[a] The cross-linked peptide was isolated as described in Materials and Methods. The ratio of the major to minor sequence was 2.3. [b] On the basis of the type C1/C2 hnRNP sequence (Swanson *et al.*, 1987), phenylalanine would be predicted at this position.

terminus of C/F. Together, these three sequences confirmed that residues 9–45 and 75–98 in C/F are identical to those in the type C1/C2 hnRNP proteins. This finding strongly suggests that C/F is derived via limited proteolysis of the type C1/C2 hnRNP proteins. Since no sequence was obtained that would correspond to cleavage after methionine 104 (as predicted by the type C1/C2 sequence), the COOH-terminus of C/F must end at or before methionine 104 in the type C1/C2 sequence. Since the 13 amino acid insert that differentiates the C1 from the C2 hnRNP protein begins at residue 107 (Burd *et al.*, 1989; Merrill *et al.*, 1989), C/F could be a proteolytic fragment of either or both of the type C1/C2 proteins. Although additional studies will be required to absolutely define the COOH-terminus of C/F, the best fit to the LDMS observed mass of 10133.7 would be for a type C1/C2 fragment that spanned residues 9–102. On the basis of the human type C1/C2 sequence (Burd *et al.*, 1989), this proposed fragment would have a predicted mass of 10 080.6. Since the ±0.52% error between this predicted and observed mass is sightly larger than our average error of ±0.2%, the amino acid sequence of calf thymus C/F may differ at one or more positions from that of the corresponding stretch of sequence in the human type C1/C2 hnRNP protein. As expected, the amino acid composition of C/F is in agreement with that predicted for residues 9–102 in the human type C1/C2 hnRNP proteins (data not shown).

Identification of the Site of C/F Photochemical Cross-Linking to $[d(T)]_8$. In preparation for these studies, 6 nmol of C/F was cross-linked to an excess of $[d(T)]_8$ as described in Materials and Methods. Approximately 25% of the input C/F was found to be cross-linked to $[d(T)]_8$. The cross-linked protein was recovered by precipitation with TCA. This step removed the majority of the free $[d(T)]_8$ but did not remove non-cross-linked protein. Since the DEAE-Sephadex column selectively retained $[d(T)]_8$ cross-linked tryptic peptides, it was not necessary to separate the cross-linked from non-cross-linked protein. The major radioactivity-containing peak from the DEAE-Sephadex column eluted with 0.5 M TEAA buffer. No such peak was detected when a non-cross-linked sample was processed through an identical protocol (data not shown).

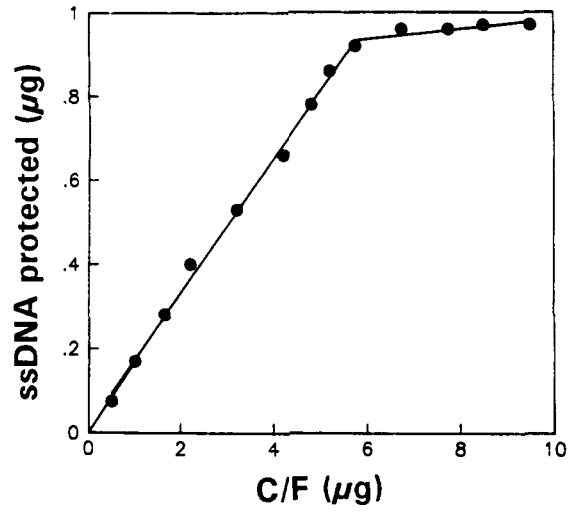

FIGURE 3: Binding stoichiometry of C/F to DNA. Measurement of the binding stoichiometry was carried out with [^{32}P]-labeled M13mp18 ssDNA as described in Materials and Methods. Each point in the figure represents the amount of ssDNA protected by C/F from nuclease digestion.

When the radioactive fractions from DEAE-Sephadex were further processed on C-4 or C-18 matrices, distinct absorbance peaks emerged that were radioactive. However, none of these peaks yielded an amino acid sequence. Most likely they represent UV-irradiated byproducts of the radiolabeled $(dT)_8$. We therefore sequenced the DEAE-Sephadex-purified radioactivity peak without further purification. As shown in Table 2, this peak contained two overlapping peptides. The major component corresponded to residues 13–29 and the minor species corresponded to residues 18–29 in the C1/C2 hnRNP. In each case, the cycle corresponding to phenylalanine 19 in the C1/C2 protein yielded no detectable amino acid residue. This result fixes the site of cross-linking at this position.

Development of a Rapid Nucleic Acid Binding Assay. Although we initially used a C/F assay based on UV-mediated cross-linking of radiolabeled oligonucleotides to the protein followed by SDS–PAGE and autoradiography, we found this approach to be too time-consuming when a number of fractions need to be simultaneously assayed. As a result, a simple TCA precipitation assay was used to quantitate the $[^{32}P][d(T)]_8$:C/F cross-linked complex. The results of this assay were compared to those obtained by SDS–PAGE. Both assay procedures indicated that UV-mediated cross-linking of $[d(T)]_8$ to C/F is linear up to at least 15 min of UV exposure (data not shown). The $[d(T)]_8$ cross-linking assay was also linear with respect to protein amounts up to at least 80 pmol (data not shown). It is clear from these data that the two assay procedures provide comparable results. The advantage of the acid precipitation assay is that a larger number of samples can be easily processed in a single day.

Nucleic Acid Binding Properties of C/F As Assayed via a UV Cross-Linking Assay. From our preliminary chromatographic studies, it was clear that C/F has high affinity for oligothymidylate residues. An examination of the ability of C/F to cross-link to $[d(T)]_x$ oligomers of differing lengths demonstrated that a minimum length of 4 was necessary for successful binding and cross-linking whereas optimal crosslinking was noted with 7–8-mers (data not shown). These data are in good agreement with the results of a nuclease protection assay from which the occluded site size of C/F on [^{3}H]-labeled M13 DNA was estimated to be 6–7 nucleotides in length. Since we have shown that the cross-linking efficiency of C/F

Table 3: Specificity of C/F Binding to Nucleic Acids

nucleic acid competitor	% cross-linking to [^{32}P][d(T)]$_8$ relative to the control[a] 1:1 ratio of competitor	5:1 ratio of competitor
d(TMP) or TpT	not determined	100
[r(A)]$_9$	98	not determined
[d(T)]$_8$	54	25
[d(T)]$_{15}$	45	11
[d(A)]^{15}	77	54
[d(C)]$_{15}$	98	67
[d(G)]$_{15}$	78	53
DNA, calf thymus	86	82
total thymus RNA	100	75
oligo(U)	23	not determined
poly(U)	28	1
poly(C)	100	100
poly(dT)	55	not determined
tRNA, *E. coli*	100	80

[a] The ability of several different nucleic acids to compete for C/F binding was determined by their ability to depress the extent of C/F photochemical cross-linking to [^{32}P][d(T)]$_8$. The competitor nucleic acid was added to the reaction mixture, which contained 26 pmol [d(T)]$_8$, prior to adding the C/F protein. C/F (56 pmols) was then added to the reaction mixtures after which the samples were incubated at 37 °C for 1 min. The samples were then chilled on ice and UV irradiated. The extent of cross-linking was then determined by TCA precipitation as described in Materials and Methods.

is ca. 25%, which was determined under saturating conditions (data not shown), it is possible to convert the fraction of C/F that is cross-linked to the fraction of C/F that is bound to [d(T)]$_8$. On the basis of this analysis we estimate that C/F has an affinity of $\sim 10^6$ M^{-1} for [d(T)]$_8$. C/F binding to [d(T)]$_8$ was not affected by 1 mM EDTA or 300 mM NaCl. The latter is in agreement with the tight binding of C/F to oligo[d(T)]-cellulose. Increasing the NaCl concentration to 500 mM or adding 1 mM iodoacetamide decreased the apparent affinity to about 8×10^5 M^{-1} while the addition of either 5 mM Mg^{2+} or 1 mM Mn^{2+} decreased the affinity for [d(T)]$_8$ to about 6×10^5 M^{-1}. As shown in Table 3, a competition assay was used to qualitatively judge the relative binding affinities of C/F for a variety of different nucleic acids. On the basis of this assay C/F has the following relative affinities for oligonucleotides: [d(T)]$_{15}$ > [d(G)]$_{15}$ ~ [d(A)]$_{15}$ > [d(C)]$_{15}$. Of the polynucleotides that were tested, C/F binds most tightly to poly(U) followed by poly[d(T)] and no significant binding was detected to poly(C). As evidenced by the relatively poor ability of *E. coli* tRNA and calf thymus DNA or RNA to compete with [d(T)]$_8$ for C/F binding, C/F binds most tightly to single-stranded ribonucleic acids.

Fluorescence Titrations. The estimated C/F binding site size of 6–8 as derived from nuclease protection and photo-cross-linking assays is in good agreement with the occluded site size of 7 ± 1 as determined from fluorescence studies of C/F binding to poly[r(ϵA)] under stoichiometric conditions (data not shown). Furthermore, the estimated affinity of 10^6 M^{-1} for C/F binding to [d(T)]$_8$ (as determined by cross-linking) is very close to the K_{app} value of 3.6×10^6 determined for C/F binding to poly[r(ϵA)] under equilibrium conditions (Table 4). To quantitatively compare preferential affinities of C/F for ssRNA versus ssDNA, the apparent association constants of C/F for poly[r(ϵA)] vs poly[d(ϵA)] were measured from the binding isotherms that are shown in Figure 4. On the basis of these data C/F has a 30-fold preferential affinity for ssRNA ($K_{app} = 3.6 \times 10^6$ M^{-1}) over ssDNA ($K_{app} = 1.2 \times 10^5$ M^{-1}). The electrostatic contribution to the free energy of binding for the C/F:poly[r(ϵA)] interaction was estimated

Table 4: Summary of C/F Binding to Various Polynucleotides in 10 mM NaCl

polynucleotide	K_{app} (M^{-1})[a]	polynucleotide	K_{app} (M^{-1})[a]
poly[r(ϵA)]	3.6×10^6	poly[r(A)]	1.1×10^6
poly[d(ϵA)]	1.2×10^5	poly[r(C)]	3.1×10^4
poly[r(G)]	6.0×10^6	ϕX ssDNA	2.0×10^6
poly[r(U)]	4.6×10^6	ϕX dsDNA	6.4×10^5

[a] K_{app} values for poly[r(ϵA)] and poly[d(ϵA)] were determined by curve fitting. Affinities for nonethenylated polynucleotides were estimated using competition assays as described under Materials and Methods.

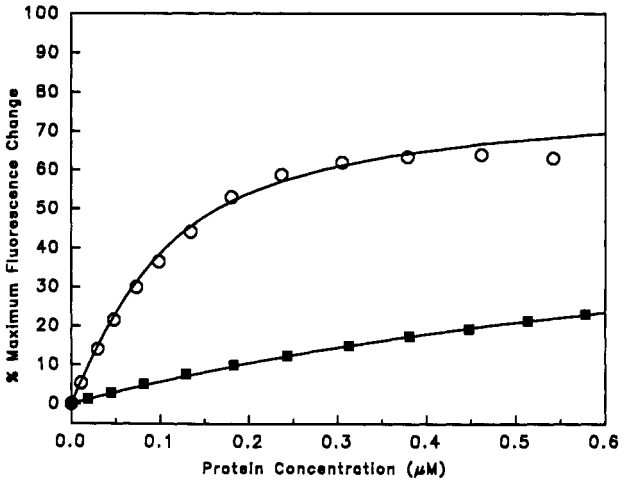

FIGURE 4: Fluorescence titrations of poly[r(ϵA)] and poly[d(ϵA)] with C/F. Starting polynucleotide concentrations were 1 μM in phosphate. Titrations were carried out in 20 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.1 mM DTT, 10% glycerol. The solid lines are theoretical best-fit curves generated as described in Materials and Methods. Open circles, C/F + poly[r(ϵA)]; filled squares, C/F + poly[d(ϵA)].

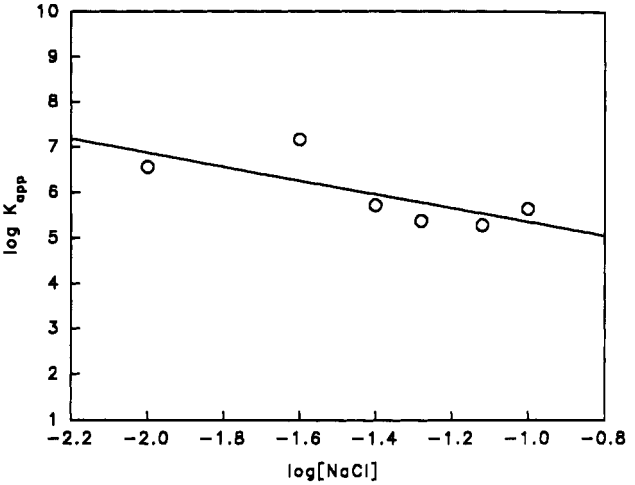

FIGURE 5: Effect of NaCl concentration on the affinity of C/F for poly[r(ϵA)]. K_{app} values were determined from titrations carried out in 20 mM Tris, pH 8.0, 0.1 mM DTT, 10% glycerol, with varying NaCl concentrations. The slope and y-intercept from the log–log plot are −1.5 and 3.8, respectively.

by measuring the affinities of C/F to poly[r(ϵA)] at several NaCl concentrations. According to Record *et al.* (1976) and Lohman *et al.* (1980), there is a linear relationship between a plot of log K_{app} versus log [NaCl] for the binding of ligands to polynucleotides. It was demonstrated in these studies that a slope of 0.7 from this "log–log" plot represents approximately one ionic interaction. Thus, the −1.5 slope of the log–log plot of C/F binding to poly[r)ϵA)] (Figure 5) suggests that there is a maximum of two ionic interactions involved in this complex.

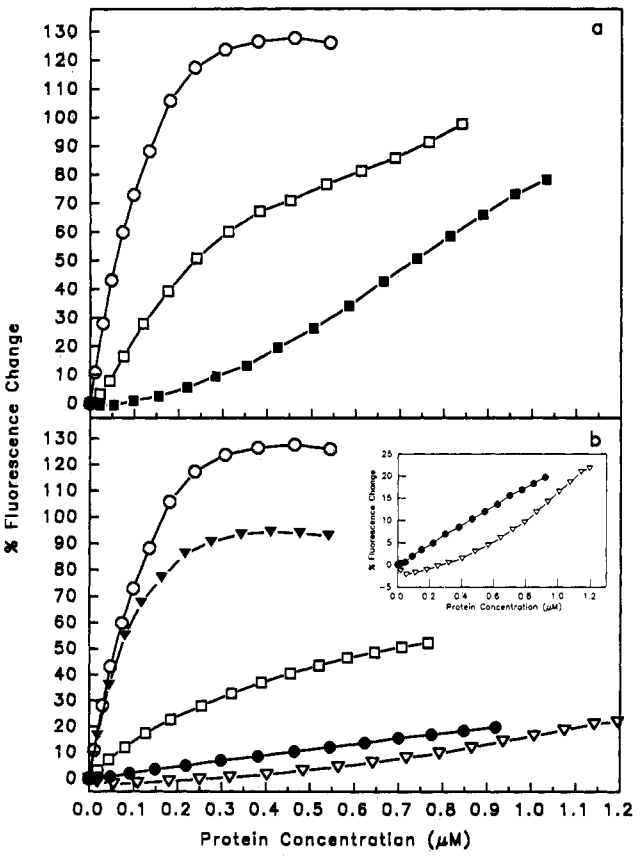

FIGURE 6: Fluorescence titrations of 1 μM poly[r(εA)] with C/F in the presence of excess competing polynucleotides using 20 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.1 mM DTT, and 10% glycerol: (a) open circles, C/F + poly[r(εA)] only; open squares, C/F + poly-[r(εA)] + 25 μM *φX174* dsDNA; filled squares, C/F + poly[r(εA)] + 25 μM *φX174* ssDNA; (b) open circles, C/F + poly [r(εA)] only; closed triangles, C/F + poly[r(εA)] + 50 μM poly[r(C)]; open squares, C/F + poly[r(εA)] + 50 μM poly[r(A)]; closed circles, C/F + poly-[r(εA)] + 50 μM poly[r(U)]; open triangles, C/F + poly[r(εA)] + 50 μM poly[r(G)]. Inset: Enlargement of titrations for poly[r(G)] and poly[r(U)] from Figure 6b showing the sigmoidal shape of the poly[r(G)] titration curve.

The relative base specificity of C/F binding to single- versus double-stranded homopolynucleotides was determined as described in Materials and Methods via a fluorescence competition assay. As summarized in Table 4, the preferential order of binding to homopolymeric RNA is poly[r(G)] ≈ poly[r(U)] > poly[r(A)] > poly[r(C)]. This order is similar to that reported previously by Swanson and Dreyfuss (1988). The only difference is that the competition data shown in Figure 6 indicate that poly[r(G)] and poly[r(U)] have approximately equal affinities whereas the previous reference found that native type C1/C2 protein bound most tightly to poly[r(U)].

As shown in Figure 6 and by the apparent association constants summarized in Table 4, C/F binds at least 3-fold more tightly to single than to double-stranded φX174 DNA. This result is consistent with our cross-linking competition assay which showed that *E. coli* tRNA and calf thymus DNA compete poorly with $[d(T)]_8$ for C/F binding. As in the case of the poly(G) titration (see inset to Figure 6b), the double-stranded φX174 titration appeared slightly sigmoidal. The latter would seem to indicate some limited cooperativity in C/F binding to poly[r(εA)] when there is a high concentration of competing ligand. Since this phenomenon was not seen in the absence of competing ligand (see Figure 4), its mechanism is unclear.

DISCUSSION

The purification of C/F was challenging due to its irreversible binding to oligo[d(T)]-cellulose, poly(U)-agarose, and C-18 reverse-phase HPLC supports. Efforts to develop a purification strategy for C/F were aided by use of the $[d(T)]_8$ cross-linking assay which enabled the rapid screening of numerous column fractions. As many, perhaps most, protein: nucleic acid complexes can be cross-linked by exposure to low doses of UV light (Williams & Konigsberg, 1992), the simplified assay procedure used for C/F would undoubtedly be applicable to many other nucleic acid binding proteins. The identity of the C/F protein as a fragment of the type C1/C2 hnRNP protein became known to us only recently. At an earlier stage of our studies, we had considered the possibility that C/F might be related to the type C1/C2 proteins because both proteins bind preferentially to poly[d(T)] and poly[r(U)]. However, the fact that type C1/C2 antibody failed to immunoprecipitate the C/F protein in either its free or oligo-[d(T)] bound form led us to believe that C/F was not related to the type C1/C2 proteins. Comparative physicochemical and, whenever possible, immunological studies suggested that C/F was not related to A1 or its UP1 degradation product (Kumar *et al.*, 1986), HMG 14 or HMG 17 (Einck & Bustin, 1985), H1 histone, a rat liver nucleoprotein named C/EBP (Johnson *et al.*, 1987), or murine leukemia virus RNA binding proteins p10 and p12. Identification of the C/F protein as being derived from the NH_2-terminal one-third of the type C1/C2 hnRNP proteins was only achieved following database searching of the amino acid sequences of cyanogen bromide peptides derived from C/F. The failure of type C1/C2 antibody to recognize C/F indicates that this antibody is directed towards the COOH-terminal two-thirds of the type C1/C2 hnRNP proteins and that the ~90-residue conserved RNA binding domain (RNP motif) in these proteins is poorly antigenic.

On the basis of amino acid sequencing and laser desorption mass spectrometry studies, calf thymus C/F spans a ~94-residue stretch of sequence that begins at threonine 9 and that probably ends at alanine 102 or glutamic acid 103 in the corresponding human type C1/C2 hnRNP sequence. Since cDNA clones have not been reported for less than full length type C1/C2 hnRNP, we presume that C/F results from limited proteolysis of intact type C1/C2 proteins. In support of this idea, we note that Dreyfuss *et al.* (1984) have previously demonstrated that both the C1 and C2 proteins can be readily converted to the identical (as judged by SDS–PAGE) ~10 000 Da fragment by limited proteolysis *in vitro*. The finding that *in vivo* photochemical cross-linking studies readily detect the type C1/C2 proteins, but not any proteins that are less than ~30 000 Da in size, as being bound to poly(A)-containing hnRNA (Dreyfuss *et al.*, 1984) also supports the notion that C/F results from limited cleavage of one or both of the type C1/C2 proteins that occurs during cell breakage and subsequent protein purification. The finding that the C/F fragment lacks the first eight residues in the type C1/C2 hnRNP protein is consistent with the observation that the first 13 residues in this fragment are disordered in solution. Hence, it is not surprising that this region is susceptible to proteolysis. The 32 000 Da A1 hnRNP and its 22 000 Da UP1 proteolytic fragment provide a parallel in regard to limited proteolysis of the type C1/C2 proteins. Since Western blots reveal only trace amounts of UP1 compared to other hnRNP-sized proteins in HeLa cells (Valentini *et al.*, 1985), it has been assumed that most UP1 that is isolated is actually produced *in vitro* by proteolysis of A1 during its purification.

Table 5: Binding Affinities for the Interaction of Individual RBD Domains with Nucleic Acids

RBD domain[a]	nucleic acid	[salt] (mM)	K_{app} (M^{-1})	reference
A1 hnRNP-II	poly[r(ϵA)]	0	4.5×10^4	Shamoo *et al.* (1993)
nucleolin-I	pre-rRNA (nonspecific)	200	$<1 \times 10^5$	Ghisolfi *et al.* (1992)
nucleolin-I	pre-rRNA (specific)	200	2.0×10^6	Ghisolfi *et al.* (1992)
C1/C2 hnRNP	poly[r(ϵA)]	10	3.6×10^6	this study
PABP-I	$(A)_{14}$	100	1.8×10^7	Sachs *et al.* (1987)
U1 A	U1 stem-loop II	50–100	1.0×10^8	Jessen *et al.* (1991)

[a] For those proteins that contain multiple RBDs, the Roman numeral indicates to which RBD the affinity refers.

As in the case of the type C1/C2 proteins, purified A1 is readily converted by limited proteolysis to a UP1-sized polypeptide that contains residues 1–185 in A1 (Kumar & Szer, 1986) and that consists almost entirely of two tandemly arranged RNA binding domains (RBD) analogous to the single RBD in C/F. These results are thus in agreement with the finding that the RBD exists in solution as a highly ordered structure consisting of four antiparallel β-strands and two α-helices (Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Wittekind *et al.*, 1992).

That the binding surface of the type A and C hnRNP is indeed very similar is suggested by photochemical cross-linking of the C/F:oligo[d(T)]$_8$ complex. This approach has previously been used to identify contact sites in the A1:oligo[d(T)]$_8$ complex (Merrill *et al.*, 1988). This study found that 75% of the cross-linking was to the NH_2-terminal RBD in A1 with the remaining 25% occurring at the second RBD. The cross-linking sites within the NH_2-terminal domain were at phenylalanine 16 (63%) and phenylalanine 58 (11.9%). On the basis of the structures of the U1A (Nagai *et al.*, 1990; Hoffman *et al.*, 1991) and type C1/C2 RBDs (Wittekind *et al.*, 1992), these two phenylalanine residues are located next to each other in two antiparallel β-strands. In the case of the isolated type C1/C2 RBD, we found a single oligo[d(T)]$_8$ cross-linking site (phenylalanine 19) which corresponds to phenylalanine 16 in the A1 RBD. Thus, even though the overall extent of direct sequence homology between the type A1 and C RBDs is only 17% both proteins appear to share a topologically similar RBD:oligonucleotide interface. By analogy, the RBD:oligonucleotide interface also does not appear to be significantly altered by incorporating the RBD domain within the context of a multidomain protein like the A1 hnRNP. The latter finding supports the view that hnRNP proteins have a modular structure (Dreyfuss *et al.*, 1993). As expected for a residue that is critically involved in nucleic acid binding, the position corresponding to phenylalanine 19 in the type C1/C2 RBD is almost exclusively occupied in other RBDs by phenylalanine, tyrosine, or another hydrophobic amino acid (Dreyfuss *et al.*, 1988; Bandziulis *et al.*, 1989; Kenan *et al.*, 1991).

Since the equilibrium binding properties of C/F appear to parallel previous results obtained on A1 fragments corresponding to one or two of its RBDs (Nadler *et al.*, 1991; Shamoo *et al.*, 1993), a general picture is beginning to emerge from these studies of the non-sequence-specific binding properties of the eukaryotic RBD. Hence, both the C/F and the 1–93 A1 RBD have an occluded site size of 6 → 7 nucleotides and both bind non-cooperatively to nucleic acids. Indeed, gel filtration studies indicate that C/F is a monomer in solution, and hence, the isolated type C1/C2 RBD does not appear to have the ability to participate in homologous RBD:RBD protein:protein interactions either when it is free in solution or bound to a polynucleotide lattice. In general, it appears that when RBD-containing proteins have the ability to participate in protein:protein interactions, the determinants for these interactions lie outside the RBD itself. For instance, in the case of A1, the ability of this protein to undergo indefinite aggregation in solution and to bind cooperatively to nucleic acids is lost upon going to a fragment containing only the two, tandemly arranged RBDs (Nadler *et al.*, 1991; Shamoo *et al.*, 1993). Similarly, since the C1 and C2 hnRNP proteins differ only by a 13-amino acid insert that follows glycine 106 in the C2 protein (Merrill *et al.*, 1989; Burd *et al.*, 1989), this insert must be involved in the ability of the type C1/C2 proteins to form a heterotetramer with the composition of $(C1)_3C2$. Since the C/F fragment ends before this insertion, it is not surprising that this fragment is a monomer in solution. Although the COOH-terminal domains of both the type A1 and C hnRNP proteins appear to be directly involved in protein:protein interactions, and thus to share at least one function in common, they nonetheless do not appear to share any significant sequence homology.

On the basis of the relative affinities for single- versus double-stranded ϕX174 DNA, both the A1 (Nadler *et al.*, 1991) and type C1/C2 RBD (Table 4) have greater preference for binding single-stranded as opposed to double-stranded nucleic acids. This differential affinity accounts for the helix-destabilizing ability of both these RBDs (Nadler *et al.*, 1991, and data not shown). Both the A1 and type C1/C2 RBD can also discriminate between single-stranded RNA versus DNA. Hence, the A1 RBD corresponding to residues 93–184 has about an 8-fold (see companion paper) and the type C1/C2 RBD has about a 30-fold (Table IV) preference for binding ssRNA over ssDNA. Although the RBD has so far only been found in RNA binding proteins, its comparatively weak preference for binding RNA over DNA raises the possibility it may also be found in proteins that bind DNA. In this regard, it is interesting that the C/F protein actually binds more tightly to double-stranded ϕX174 DNA than to poly[r(C)] and its affinity for poly[r(A)] is only about 2-fold above that for double-stranded ϕX174 DNA (Table 4).

When taken together with the preceding study by Shamoo *et al.*, our results indicate that the affinity of the C/F fragment for poly[r(ϵA)] is at least 100-fold higher than that for either of the corresponding type A1 RBDs (residues 1–92 or 93–184) and about 10-fold higher than that for the intact 1–184 A1 domain which spans both RBDs of A1. One factor which partially accounts for the higher affinity of C/F for RNA is its higher net positive charge. Hence, while the type A RBDs have net charges (at pH 8) of ca. 0 and 1– for 1–92 and 93–184, respectively, the C/F fragment has a net positive charge of ca. 8+. The significant role played by electrostatic interactions in the tight-binding of C/F to poly[r(ϵA)] is indicated by the fact that the slope of the log affinity versus log [NaCl] plot for the C/F:poly[r(ϵA)] interaction (at pH 8) is approximately twice that of the corresponding A1 93–184 RBD domain (at pH 6). On the basis of these data, binding of C/F involves a maximum of two ionic interactions as opposed to the single interaction for the A1 93–184 domain. By extrapolating to 1 M NaCl, we estimate that up to 46%

of the free energy of binding of the C/F protein to poly-[r(εA)] derives from electrostatic interactions as opposed to 27% in the case of the A1 93–184 RBD.

A recent multidimensional NMR study of the type C1/C2 RBD:oligo[r(U)] complex may offer additional insight into the molecular basis for the higher affinity of the type C1/C2 as opposed to type A RBD. This study (Görlach *et al.*, 1992) showed that while the overall structure of the RBD did not change upon binding $[r(U)]_8$, resonances originating from several amino acid residues underwent unusually large chemical shifts in the presence of the oligonucleotide. Since most of these residues are located either in the β-sheet "platform" that is thought to represent the RBD binding surface of the type C1/C2 RBD or are contained within the contiguous NH_2- and COOH-terminal regions, many of these residues are almost certainly at the interface of the type C1/C2 RBD:oligo$[r(U)]_8$ complex. It is interesting that of the 10 type C1/C2 residues that underwent the largest chemical shifts (Thr-9, Asp-10, Ser-13, Asn-15, Ser-16, Val-18, Ile-20, Gln-56, Asn-83, and Ala-86), only Ile-20 is found at the same position in the A1 RBD-I when the two RBDs are aligned on the basis of their secondary structures (Garrett *et al.*, 1993). It would be interesting therefore to use *in vitro* mutagenesis to place some of these type C1/C2 RBD residues in the corresponding positions in the A1 RBD-I and to determine if the resulting hybrid RBD began to acquire the higher binding affinities of the type C1/C2 RBD. In this regard it would be particularly interesting to substitute the glutamic acid that is found at position 10 in the A1 RBD with the serine (*i.e.*, residue 13) that is found at the corresponding position (when the A1 RBD-I and type C1/C2 RBD are aligned on the basis of their secondary structures (Garrett *et al.*, 1993) in the type C1/C2 RBD. At least in the case of the U1A protein, several hydrogen-bonding interactions (that were mediated via Thr, Asn, and Tyr residues) were found to play a critical role in binding to U1 RNA (Jessen *et al.*, 1991).

Although the affinity of the type C1/C2 RBD for single-stranded RNA appears to be about 100-fold higher than that of either of the type A RBDs, it is nonetheless considerably lower than that for more specific RBD:RNA interactions. As shown in Table 5, the affinity of the type C1/C2 RBD for poly[r(εA)] appears to be about 5-fold less than the affinity of RBD-IV from the poly(A) binding protein for $(A)_{14}$ (Sachs *et al.*, 1987) and is about 30-fold less than the affinity of the RBD domain of the U1 A snRNP protein for its stem-loop 2 target in U1 RNA (Jessen *et al.*, 1991). Obviously, as the requirements for RBD binding specificity are increased, the target affinities must also be increased. Presumably, this is accomplished by subtle amino acid changes in the β-pleated sheet and in the contiguous loop regions that are thought to form the RBD binding surface (Jessen *et al.*, 1991; Kenan *et al.*, 1991). Using this line of reasoning (Kenan *et al.*, 1991; Dreyfuss *et al.*, 1993), those RBD residues that are most highly conserved are probably essential either for maintaining the overall structure of the RBD, which on the basis of studies on types A (Garrett *et al.*, 1993), C (Wittekind *et al.*, (1992), and U1A RBD (Nagai *et al.*, 1990; Hoffman *et al.*, 1991) is highly conserved, or for maintaining a basal level of nucleic acid binding affinity that is probably common to all RBDs. Specificity would thus be attained either by adding additional interacting residues, as U1A has in loop 3 (Scherly *et al.*, 1990; Bentley *et al.*, 1991; Jessen *et al.*, 1991), or by substituting more favorable amino acid residues at existing sites of contact between the RBD and its nucleic acid target.

ACKNOWLEDGMENT

We thank Dr. J. Wilusz for many useful discussions and for carrying out the C/F immunoprecipitation studies with the type C1/C2 hnRNP antibodies that were kindly provided by Dr. G. Dreyfuss. We also thank the staff of the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University, particularly Ms. Nancy Williams, for assistance with the protein chemistry studies and Dr. John Rush of the Harvard Microchemical Facility for the laser desorption mass spectrometric analyses.

REFERENCES

Bacerra, S. P., & Wilson, S. H. (1984) *Biochemistry 23*, 904–914.

Bandziulis, R. J., Swanson, M. S., & Dreyfuss, G. (1989) *Genes and Dev. 3*, 431–437.

Barnett, S. F., Friedman, D. L., & LeStourgeon, W. M. (1989) *Mol. Cell. Biol. 9*, 492–498.

Beyer, A. L., Christensen, M. E., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell 11*, 127–138.

Burd, C. G., Swanson, M. S., Görlach, M., & Dreyfuss, G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9788–9792.

Casas-Finet, J. R., Smith, J. D., Kumar, A., Kim, J. G., Wilson, S. H., & Karpel, R. L. (1993) *J. Mol. Biol. 229*, 873–889.

Chase, J. W., & Williams, K. R. (1986) *Annu. Rev. Biochem. 55*, 103–136.

Cobianchi, F. Karpel, R. L., Williams, K. R., Notario, V., & Wilson, S. H. (1988) *J. Biol. Chem. 263*, 1063–1071.

Dreyfuss, G., Choi, Y. D., & Adam, S. A. (1984) *Mol. Cell. Biol. 4*, 1104–1114.

Dreyfuss, G., Swanson, M. S., & Pinol-Roma, S. (1988) *Trends Biochem. Sci. 13*, 86–91.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., & Burd, C. G. (1993) *Annu. Rev. Biochem. 62*, 289–321.

Edenberg, H. J., Anderson, S., & DePamphilis, M. L. (1978) *J. Biol. Chem. 253*, 3273–3280.

Einck, L., & Bustin, M. (1985) *Exp. Cell Res. 156*, 295–310.

Garrett, D. S., Lodi, P., Shamoo, Y., Williams, K. R., Clore, G. M., & Gronenborn, A. M. (1993) *Biochemistry*, submitted.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., & Dreyfuss, G. (1992) *EMBO J. 9*, 3289–3295.

Hoffman, D. W., Query, C. Q., Golden, B. L., White, S. W., & Keene, J. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495–2499.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447–3456.

Kenan, D. J., Query, C. C., & Keene, J. D. (1991) *TIBS 16*, 214–220.

Kumar, A., Williams, K. R., & Szer, W. (1986) *J. Biol. Chem. 261*, 11266–11273.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

LeStourgeon, W. M., Barnett, S. F., & Northington, S. J. (1990) in *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies* (Strauss, P., & Wilson, S., Eds.) Vol. 2, pp 477–502, Telford Press, New Jersey.

Merrill, B. M., & Williams, K. R. (1990) in *The Eukaryotic Nucleus, Molecular Biochemistry and Macromolecular Assemblies* (Strauss, P. R., & Wilson, S. H., Eds.) pp 579–604.

Merrill, B., Stone, K., Cobianchi, F., Wilson, S., & Williams, K. (1988) *J. Biol. Chem. 263*, 3307–3313.

Merrill, B. M., Barnett, S. F., Le Stourgeon, W. M., & Williams, K. R. (1989) *Nucleic Acids Res. 17*, 8441–8449.

Miller, J. (1972) in *Experiments in Molecular Genetics*, pp 431–432, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Modak, M. J. (1978) *Biochemistry 17*, 3116–3120.

Modak, M. J. (1979) *Biochemistry 18*, 2679–2684.

Modak, M. J., & Gillerman-Cox, E. (1982) *J. Biol. Chem. 257*, 15105–15109.

Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., & Williams, K. R. (1991) *Biochemistry 30*, 2968–2976.
Nagai, K., Oubridge, C., Jessen, T., Li, J., & Evans, P. (1990) *Nature 348*, 515–520.
O'Donnell, M. E., Elias, P. Funnel, B. E., & Lehman, I. R. (1987) *J. Biol. Chem. 262*, 4260–4266.
Overman, L. B., Bujalowski, W., & Lohman, T. (1988) *Biochemistry 27*, 456–470.
Pandey, V. N., & Modak, M. J. (1987) *Preparative Biochem. 17*, 359–377.
Pinol-Roma, S., Choi, Y. D., Matunis, M. J., & Dreyfuss, G. (1988) *Genes Dev. 2*, 215–227.
Sachs, A. B., Davis, R. W., & Kornberg, R. D. (1987) *Mol. Cell. Biol. 7*, 3268–3276.
Schreier, P. H., & Cortese, R. (1979) *J. Mol. Biol. 129*, 169–172.
Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205–280.
Swanson, M. S., & Dreyfuss, G. (1988) *Mol. Cell. Biol. 8*, 2237–2241.
Valentini, O., Biamonti, G., Pandolfo, M., Morandi, C., & Riva, S. (1985) *Nucleic Acids Res. 13*, 337–346.
Wilk, H., Angeli, G., & Schäfer, K. (1983) *Biochemistry 22*, 4592–4600.
Williams, K. R., & Konigsberg, W. (1992) *Methods Enzymol. 208*, 516–539.
Wilson, S. H., Cobianchi, F., & Guy, H. R. (1987) in *DNA: Protein Interactions and Gene Regulation* (Thompson, E. B., & Papaconstantinou, J., Eds.) pp 129–146, University of Texas Press, Austin, TX.
Wilusz, J. D., Feig, L., & Shenk, T. (1988) *Mol. Cell. Biol. 8*, 4477–4483.
Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G., & Mueller, L. (1992) *Biochemistry 31*, 6254–6265.
Zamore, P. D., Patton, J. G., & Green, M. R. (1992) *Nature 355*, 609–614.