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[†]

James G. McAfee, Syrus R. Soltaninassab, Mark E. Lindsay, and Wallace M. LeStourgeon*

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

Received August 21, 1995; Revised Manuscript Received November 27, 1995[⊗]

ABSTRACT: Proteins C1 and C2 together comprise about one-third the protein mass of mammalian core 40S heterogeneous nuclear ribonucleoprotein particles (40S hnRNP) and exist as heterotetramers of $(C1)_3C2$. On the basis of nonequilibrium binding studies, it has been suggested that the C proteins specifically bind oligo(U)- and poly(U)-rich sequences, and preferentially associate with uridine-rich regions near the 3′ termini of many introns. We describe here a more quantitative characterization of the equilibrium binding properties of native and recombinant C protein to homoribopolymers using fluorescence spectroscopy. Like C protein from HeLa cells, the recombinant proteins spontaneously oligomerize to form tetramers with the same hydrodynamic properties as native protein. Near-stoichiometric binding titrations of the fluorescent homoribopolymer polyethenoadenosine (poly[r(ϵA)]) with recombinant $(C1)_4$ and $(C2)_4$ homotetramers along with competition binding assays with poly(A) and poly(C) indicate that the binding site size (n) is between 150 and 230 nucleotides. This site size range is in close agreement with that previously determined for native C protein through hydrodynamic and ultrastructural studies (∼230 nucleotides). $(C1)_4$ and $(C2)_4$ bind poly(G) with intrinsic affinities (K_i) of 10^9 M^{-1}, which are a hundredfold higher than their affinities for poly(U). In opposition to reports that C protein does not bind poly(A) and poly(C), we find that the C proteins bind these substrates with moderate K_i, but with high cooperativity (ω). The overall affinity ($K\omega$) for the binding of both proteins to poly(A) and poly(C) is 10-fold higher ($>10^8$ but $<10^9$ M^{-1}) than their affinities for poly(U). The highly cooperative binding of C protein to these substrates provides a mechanistic basis for the distribution of C protein along the length of nucleic acid substrates.

Many observations indicate that during transcription in higher eucaryotes the elongating pre-mRNA is bound by a unique set of abundant nuclear proteins to form a repeating array of 20−22 nm particles termed 30−40S heterogeneous nuclear ribonucleoprotein particles (hnRNP particles or ribonucleosomes)[1] (Wooley *et al.*, 1986). For example, when HeLa nuclear lysates are prepared under conditions where nuclease activity is not aggressively inhibited, most of the pre-mRNA is recovered from the 30−40S region of density gradients. Electron micrographs of this material reveal spherical particles with diameters of 20−22 nm (termed monoparticles). On the other hand, when nuclear lysates are prepared under conditions of nuclease inhibition, the pre-mRNA sediments throughout the 30−300S region of gradients. Ultrastructural analysis of this material reveals monoparticles at 30−40S, dimers at 60−80S, and oligomers of 22 nm particles in the faster sedimenting regions (Samarina & Krichevskaya, 1981; Sperling & Sperling, 1990; LeStourgeon, *et al.*, 1990; Dreyfuss, 1986).

Monoparticles isolated from density gradients consist primarily of six abundant nuclear proteins known as the core particle proteins (designated A1, A2, B1, B2, C1, and C2) (Beyer *et al.*, 1977). When monoparticles or polyparticles are isolated from actively growing HeLa cells in the absence of protease or extensive nuclease activity, these proteins are recovered in the approximate molar ratio of 3A1:1B2:3A2:1B1:3C1:1C2. This unique stoichiometry appears to result from the association of A2 with B1 and C1 with C2 to form heterotetramers (Barnett *et al.*, 1989, 1991; LeStourgeon *et al.*, 1990; Huang *et al.*, 1994). Evidence suggesting that A1 also exists in an oligomeric state is seen in the recovery of A1 trimers from cross-linked monoparticles (Lothstein *et al.*, 1985; Harris *et al.*, 1988a,b). The $(A2)_3B1$ and $(C1)_3C2$ tetramers assemble from the product of alternatively spliced transcripts (Merrill *et al.*, 1989; Burd *et al.*, 1989). For example, protein C2 differs from C1 by the presence of a 13 amino acid insert (at Gly 106) in the 290 amino acid sequence of C1. As demonstrated in this report, a physical characterization of recombinant C1 and C2 eliminates the possibility that the 13 amino acid insert functions in oligomerization or tetramer stabilization.

If isolated hnRNP complexes are digested to completion with nuclease, the core particle proteins dissociate, resulting

[†] This work was supported by the National Institutes of Health (GM48567) (W.M.L.) and by a Ford Foundation Postdoctoral Fellowship (880-0765) (J.G.M.). M.E.L. was the recipient of a Howard Hughes Biomedical Foundation grant for undergraduate science research.

* Address correspondence to this author at the Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235. Telephone: 615-322-2588. FAX: 615-343-6707. Email: lestouwm@ctrvax.vanderbilt.edu.

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1996.

[1] Abbreviations: hnRNP, heterogeneous nuclear ribonucleoprotein; poly[r(ϵA)], poly(ethenoadenosine); poly(U), polyuridine; poly(A), polyadenosine; poly(G), polyguanosine; poly(C), polycytidine; kDa, kilodalton(s); K_i, intrinsic association constant; ω, cooperativity parameter; n, occluded binding site size; oligo-U, oligouridine; SDS−PAGE, sodium dodecyl−sulfate polyacrylamide gel electrophoresis; $(C1)_4$, homotetramer of protein C1; $(C2)_4$, homotetramer of protein C2; RRM, RNA recognition motif.

0006-2960/96/0435-1212$12.00/0 © 1996 American Chemical Society

in the loss of monoparticle structure. Dissociated monoparticles spontaneously reassemble upon the addition of RNA approximately 700 nucleotides in length. Reconstituted monoparticles possess the same core protein composition, ultrastructural morphology, and sedimentation characteristics as native monoparticles (Conway *et al.*, 1988; Huang *et al.*, 1994; Rech *et al.*, 1995a; Wilk *et al.*, 1983). A maximum of 2 and 3 monoparticles reconstitute on RNA lengths of approximately 1400 and 2100 nucleotides, respectively. These studies demonstrate that formation of the monoparticle is a reversible pathway that can be observed and studied *in vitro*. This laboratory has recently isolated and characterized an obligate first-step intermediate in the pathway to monoparticle assembly (Huang *et al.*, 1994; Rech *et al.*, 1995a). Ultrastructural studies show the intermediate to have a geometrically distinct triangular structure. This structure has a sedimentation coefficient of 19S, and consists of 3 C protein tetramers bound to approximately 700 nucleotides of RNA transcript (Huang *et al.*, 1994; Rech *et al.*, 1995a). This stoichiometry indicates that a single tetramer associates with a maximum of about 230 nucleotides of RNA. Ultrastructural studies have confirmed that only 1 tetramer associates with a 230-nucleotide length of RNA. The 19S C protein−RNA complex nucleates 40S monoparticle reconstitution when added to nuclease-dissociated hnRNP complexes. Since the 19S complex is a key intermediate in the formation of the 40S monoparticle, it has been an objective of this laboratory to quantitatively investigate the binding of C protein to RNA.

It was initially argued from electron micrographs of spread chromatin that the core particle proteins do not bind RNA in a sequence-specific manner because large and small transcripts appear to be packaged into contiguous arrays of uniform-sized particles (Malcolm & Sommerville, 1977; Vazquez-Nin *et al.*, 1994). Similarly, several immunofluorescent studies (Kiseleva *et al.*, 1994; Amero *et al.*, 1992; Beyer & Osheim, 1990) and biochemical studies (van Eekelen *et al.*, 1982; Lothstein *et al.*, 1985; Harris *et al.*, 1988a; Hendrick & De Kloet, 1975) have generally confirmed the uniform distribution of the core particle proteins on nascent transcripts. Also in support of nonspecific binding is the finding that hnRNP particles spontaneously assemble *in vitro* on 700 ± 20 nucleotide increments of RNA or DNA to form contiguous arrays. This occurs whether or not RNA splicing signals are present (Conway *et al.*, 1988; Huang *et al.*, 1994; Wilk *et al.*, 1983). Several reports, however, suggest that the C proteins bind RNA in a sequence-specific fashion (Swanson & Dreyfuss, 1988a,b; Hamilton *et al.*, 1993; Olsen *et al.*, 1992; Wilusz & Shenk, 1990; Gorlach *et al.*, 1994). For example, C protein from HeLa nucleoplasm was shown to form a complex with poly(U) that did not dissociate at 2 M NaCl (Swanson & Dreyfuss, 1988a). It was subsequently reported that C protein binds specifically to U-rich regions often present near the 3′ end of introns, and that C protein does not bind to bacterial RNA or RNA lacking introns (Swanson & Dreyfuss, 1988b). In addition, recombinant C1 has been used to selectively isolate high-affinity sequences from a randomized pool of RNA sequences (SELEX procedure) (Gorlach *et al.*, 1994). This study showed that recombinant C1 exhibited preferential affinity for sequences containing at least five contiguous uridylates. These studies infer that C protein functions in some capacity during RNA splicing by specifically binding polypyrimidine tracts near the 3′ termini of introns. This inference has been supported by the report that antibodies against C protein block RNA splicing *in vitro* (Choi *et al.*, 1986). However, it was not demonstrated that the addition of C protein to these extracts could restore activity. Antibodies against C protein could attenuate splicing through a number of nonspecific mechanisms. Furthermore, a physiologically meaningful interpretation of the SELEX data is complicated by the absence of five contiguous uridylates at the 3′ end of many introns, by the presence of this sequence in nascent transcripts at sites other than polypyrimidine tracts (Senapathy, 1986), and by the experimental design of the selection procedure which does not take into account the RNA binding site size of C protein nor cooperative protein binding events.

The state of C protein oligomerization has been neglected in many studies. This oversight is a barrier to a mechanistic understanding of C protein function in RNA packaging and processing. Sedimentation analyses, ion-exchange and size-exclusion chromatography, and protein cross-linking studies indicate that native C protein exists in solution and in intact 40S hnRNP particles as highly stable heterotetramers composed of $(C1)_3C2$ but they do not exclude the existence of $(C1)_4$ and $(C2)_4$ homotetramers (Barnett *et al.*, 1989; Huang *et al.*, 1994; Rech *et al.*, 1995a). If C protein binds RNA as a tetramer possessing 4 identical RNA binding domains with a site size greater than 200 nucleotides, it becomes difficult to conceptualize a mechanism for sequence specificity. In this report, we demonstrate that recombinant C1 and C2 spontaneously oligomerize to form highly stable homotetramers that are indistinguishable from native tetramer based on sedimentation analyses, gel exclusion chromatography, ultrastructural studies, and RNA binding site size. Since all of the studies to date characterizing the binding properties of native or recombinant C protein have been nonequilibrium in nature, we present here a quantitative investigation of the equilibrium binding properties of recombinant $(C1)_4$ and $(C2)_4$ to various RNA substrates. The recombinant proteins are preferred substrates in these equilibrium binding studies because they lack the posttranslational heterogeneity (and possible subunit heterogeneity) of the native protein.

MATERIALS AND METHODS

Generation and Expression of Recombinant C1 and C2 Clones. Using specific primers derived from the published sequence of the C protein gene (Swanson *et al.*, 1987; Burd *et al.*, 1989) (containing *Nde*1 and *Bam*H1 tags at the 5′ and 3′ ends of the gene), C1- and C2-specific cDNAs were produced by reverse transcription of cytoplasmic HeLa RNA followed by PCR amplification (Sambrook *et al.*, 1989). The specific PCR products were purified from agarose gels and cloned into the *Nde*1/*Bam*HI site of the expression vector pET-3b. This DNA construct was transformed into *E. coli* DH5α, and cells containing recombinant C1 and C2 plasmids were identified by colony hybridization followed by Southern analysis of plasmid DNA from selected clones (Southern, 1975).

Purification of Native and Recombinant C Protein. HeLa C protein tetramers were purified under native conditions as described previously (Barnett *et al.*, 1988; Huang & LeStourgeon, 1994). To express proteins C1 and C2 in bacterial cells, recombinant plasmids were transformed into

the expression strain *E. coli* Bl21(DE3) pLysS. The growth of this strain and induction of the recombinant proteins were as previously described (Studier *et al.*, 1990). For protein purification, typically 4 L of induced cells was harvested 2 h postinduction by centrifugation at 6000 rpm in an HB4 rotor for 30 min in a Sorvall RC2-B centrifuge. The cell pellet was resuspended in LB media (10 g of bactotryptone, 10 g of NaCl, and 5 g of yeast extract) and stored at −70 °C. Lysis of the cells was facilitated by thawing at 20 °C. To the thawed lysate were added phenylmethane sulfonyl chloride (PMSF) and $CaCl_2$ to a final concentration of 1 mM. Nucleic acids were then removed by digestion with micrococcal nuclease and DNase I at 37 °C for 30 min. The lysate was then centrifuged at 10 000 rpm in a Sorvall HB4 rotor for 20 min. The supernatant was then diluted 10-fold in buffer A (10 mM Tris, pH 8.0, 5 mM EDTA, 3 mM DTT), and applied to a Bio-rad Econo-Pac Q anion-exchange column. The column was washed with buffer A containing 200 mM NaCl, followed by elution with a linear gradient of 200−800 mM NaCl in the same buffer. Fractions containing recombinant C protein were combined and diluted 2-fold in buffer A. The diluted fractions were applied to a single-stranded DNA−cellulose column (Pharmacia 27-5579-02) equilibrated with TE (10 mM Tris, pH 8.0, 1 mM EDTA) containing 100 mM NaCl. After application of the sample, the column was washed with 2−3 column volumes of this buffer followed by elution of the C protein with TE buffer containing 0.9 M NaCl. Protein purified from bacterial cell lysates in this manner appears to be greater than 99% pure as judged by SDS−PAGE analysis. Purified C protein was quantified using the BCA assay (Pierce 23225). Accuracy of the BCA assay for C protein was determined through quantitative amino acid analysis. Quantitative amino acid analyses were also used to demonstrate the absence of tryptophan in the purified native and recombinant C protein preparations.

Sedimentation, Size-Exclusion Chromatography, and Electron Microscopy. The sedimentation, chromatographic, and ultrastructural properties of purified $(C1)_4$, $(C2)_4$, and native $(C1)_3C2$ tetramers were compared as follows: For the sedimentation studies, 0.5 mL of the various protein samples in pH 8 STM buffer was mixed with 0.5 mL of pH 8 STM buffer containing 50 mg of bovine serum albumin (Sigma) and immunoglobulin G (Sigma). These preparations were layered on 13 mL 5−20% linear glycerol gradients prepared in Beckman SW 40.1 centrifuge tubes. The preparations were centrifuged for 28 h at 38K rpm at 4 °C in the SW 40 rotor using a Beckman Model L3-50 centrifuge. The distribution of protein in the gradients was monitored continuously using a 0.2 mL flow cell in a Hewlett Packard Model 8252A spectrophotometer and simultaneously fractionated into 18−0.7 mL aliquots. Protein in each fraction was precipitated with three volumes of ethanol, dried under vacuum, and solublized in SDS-containing sample buffer for electrophoresis as described elsewhere (Huang & LeStourgeon, 1994).

Size-exclusion chromatography was conducted using a 2.5 × 90 cm glass column packed with Sephacryl S-300 (Pharmacia LKB Laboratories). For most experiments, the column was equilibrated and eluted with pH 8.0 STM buffer at a NaCl concentration of 400 mM. In some cases, the column was equilibrated and eluted with buffers containing NaCl concentrations of 90, 200, and 1000 mM, with buffers containing 3% Triton X-100, 20 mM EDTA, 200 mM 2-mercaptoethanol, and 10 mM ATP and/or GTP. In most cases, horse spleen ferritin or apoferritin, bovine serum albumin, and bovine liver catalase were added to the samples as internal size standards prior to sample loading. The column was monitored at 280 nm with a Gilson Model 111 spectrophotometer, and the protein in consecutive 4 mL fractions was prepared for electrophoresis as described elsewhere (Huang *et al.*, 1994).

For ultrastructural comparison of the recombinant and native tetramers, both fixed and unfixed samples were examined. Fixed samples were allowed to stand for 10 min on ice with 1% glutaraldehyde at pH 7.0. Samples were applied to Formvar−carbon-coated grids that were glow-discharged within 30 min of sample application. The grids were stained with filtered 2% uranyl acetate, blotted, and dried over a beaker of distilled water. In some experiments, ferritin was added for use as an internal size standard. Sample preparation and platinum shadowing were performed as described in detail elsewhere (Rech *et al.*, 1995a,b). In some experiments, tobacco mosaic virus or myosin was used as internal standards at a final concentration of 20−25 mg/ml. Electron microscopy was performed on an Hitachi H-800 transmission electron microscope at 100 kV or on a Phillips CM-12 transmission electron microscope at 120 kV. Micrographs were recorded at nominal magnifications of 40 000x and 70 000x. Microscope calibrations were performed by photographing tropomyosin paracrystals at appropriate magnification followed by measuring the 395 Å internal repeats. Measurements of the various preparations were made from negatives using a Nikon profile projector. On average, 150 measurements were made for each sample.

Equilibrium Binding Analyses Using Protein-Induced Enhanced Fluorescence of Poly[r(ϵA)]-Labeled RNA. Quantitative analyses of the binding of purified recombinant $(C1)_4$ and $(C2)_4$ to various RNA substrates were obtained by measuring the enhancement of the poly[r(ϵA)] fluorescence upon protein binding. Poly[r(ϵA)] was prepared by treating poly(A) with chloracetaldehyde as described by Steiner *et al.* (1973). Fluorescence measurements were carried out on an SLM Aminco Bowman Series 2 Luminescence Spectrometer, with excitation and emission slits fixed at 4 and 8 nm, respectively. Titrations were conducted by exciting a fixed amount of poly[r(ϵA)] at 310 nm, and recording the change in its emission at 410 nm as a function of increasing protein concentration. Fluorescence measurements were corrected for dilution due to the addition of protein sample, and for background fluorescence. In these experiments, it is assumed that the fractional change in the fluorescence enhancement (E) is equal to the fractional saturation of the nucleic acid lattice. Binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ to poly[r(ϵA)] were estimated by simulating theoretical curves to fit the experimental data using the McGhee−von Hippel noncooperative model (McGhee & von Hippel, 1974). In these simulations, the binding density, ν (moles of protein bound per mole of nucleotides), was varied from 0 to 1/binding site size (n) (the maximum possible binding density). Theoretical enhancement (E) values were calculated for a given set of binding parameters, using the relationship:

$$E = \nu n E_{max}$$

where E_{max} is the maximum enhancement observed when the nucleic acid lattice is saturated. The total protein concentration corresponding to calculated E values was determined from the equation:

$$L_t = L_b + L = \nu R_t + L$$

where L_t, L_b, and L are the total, bound, and free protein, respectively, while R_t is the total RNA concentration at any one point during the titration. Binding parameters determined by simulations were virtually identical to those determined by nonlinear least-squares regression analysis of the data. Binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(U), poly(C), poly(G), and poly(A) were obtained by analysis of competition titrations (see Results) based on a previously described method (Kowalczykowski *et al.*, 1986).

RESULTS

Oligomerization State of Native and Recombinant C Proteins. Chemical cross-linking, analytical ultracentrifugation, and ultrastructural studies have demonstrated that native C protein purified from isolated HeLa cell hnRNP complexes exists as $(C1)_3C2$ heterotetramers (Barnett *et al.*, 1989; Huang *et al.*, 1994; Rech *et al.*, 1995b). If homogeneous $(C1)_4$ and $(C2)_4$ tetramers exist *in vivo*, they were not detected through these analytical procedures. Purified native tetramers elute anomalously from size-exclusion columns. For example, HeLa C protein tetramers with a sequence-determined mass of 129 056 Da sediment in density gradients with an approximate mass of 130 000 Da but elute from Sephacryl S-300 columns as a homogeneous entity with an apparent mass near 400 kDa (Barnett *et al.*, 1989; Huang *et al.*, 1994). It has been suggested that the anomalous elution from size-exclusion columns may in part be due to molecular asymmetry and perhaps to an unusually large charge-induced hydration shell (Barnett *et al.*, 1989). When expressed in bacterial cells, proteins C1 and C2 spontaneously oligomerize to form $(C1)_4$ and $(C2)_4$ tetramers which elute coincident with native tetramers from Sephacryl S-300 columns with an apparent mass near 400 kDa (not shown). Also like native C protein, the recombinant forms sediment in density gradients with a mass near 130 kDa (not shown). In addition, the recombinant homotetramers have the same elution characteristics as native protein from strong anion columns (not shown). These findings indicate that oligomerization is an intrinsic property of the C proteins, that the 13 amino acid insert in C2 does not modulate oligomerization, and that recombinant forms of C1 and C2 assume a quaternary structure similar to, if not the same as, native protein.

Spectroscopic Properties of Native and Recombinant C Proteins. Native HeLa C protein contains eight tyrosine residues and lacks tryptophan (Swanson *et al.*, 1987; Burd *et al.*, 1989). The absence of tryptophan in preparations of purified native and recombinant C1 and C2 was confirmed via quantitative amino acid analysis. The emission maximum for tyrosine in proteins generally ranges from 302 to 310 nm (Ross *et al.*, 1991). Native C protein, however, has an emission maxima at 338 nm an uncharacteristic maxima for a protein containing only tyrosine and lacking tryptophan. Denaturation of the native protein with 8 M guanidine hydrochloride shifts the emission maximum to that of tyrosine at 306 nm (data not shown). These results indicate that the 338 nm emission maximum of native C protein results from the formation of tyrosinate which is believed to arise from tyrosine undergoing excited state proton transfer in the folded protein (Ross *et al.*, 1991). $(C1)_4$ and $(C2)_4$ have the same tyrosinate emission peak, indicating that the tyrosine giving rise to this fluorescence is in the same environment in both the native and recombinant species (not shown). The fluorescence emission and excitation spectra, like the hydrodynamic studies described above, suggest that $(C1)_4$ and $(C2)_4$ possess the same global folding pattern as native heterotetramer.

RNA Binding Site Size of C Protein. In vitro reconstitution studies have demonstrated that 40S monoparticle assembly is a function of RNA length (Conway *et al.*, 1988; Huang *et al.*, 1994). Monoparticles assemble on approximately 700 nucleotide lengths of RNA while dimers and trimers assemble on multiples of this length. The length of RNA recovered from isolated native 40S monoparticles ranges from 500 to 1000 nucleotides (Beyer *et al.*, 1977; Wilk *et al.*, 1983; Pullman & Martin, 1983). Previous studies have shown that C protein binds 700 nucleotide lengths of RNA *in vitro* to form a geometrically distinct 19S triangular structure that is present in native 40S monoparticles and that can nucleate *in vitro* reconstitution of 40S monoparticles (Huang *et al.*, 1994; Rech *et al.*, 1995a). The identical RNA length-dependency for assembly of the 40S monoparticle and 19S complex suggests that C protein−RNA interactions determine the length of RNA packaged in each 40S monoparticle. However, *in vitro* reconstitution studies do not rigorously address the issue of the occluded binding site size of the C protein tetramer. To determine the stoichiometry of $(C1)_4$ and $(C2)_4$ binding to poly[r(ϵA)], a titration was carried out in TE buffer containing 100 mM NaCl (Figure 1). Under these conditions of ionic strength, the highest affinity for both proteins to this RNA substrate was observed. Neither native tetramers (not shown) nor $(C1)_4$ or $(C2)_4$ homotetramers bind stoichiometrically to poly[r(ϵA)] under these conditions. An approximation to stoichiometric binding was obtained by extrapolation from the linear region of the graph (in the region of RNA excess) to an intersecting line at the equivalence point (shown by the intersection of the two straight lines in Figure 1). An estimate of 230 nucleotides was obtained for the binding site size of the $(C1)_4$ and $(C2)_4$ tetramers, respectively. Under these conditions, the native tetramer has a site size of 220 nucleotides. These numbers are essentially identical to the reported binding site size of 230 nucleotides estimated for native C protein through hydrodynamic and ultrastructural studies (Huang *et al.*, 1994).

Binding Affinities of $(C1)_4$ and $(C2)_4$. To obtain the intrinsic association constants (K_i) for the interaction of the $(C1)_4$ and $(C2)_4$ tetramers with poly[r(ϵA)] in 100 mM NaCl, theoretical curves were generated using the McGhee−von Hippel noncooperative model (Figure 2) (McGhee & von Hippel, 1974). The derived association constants are 3.9×10^6 M^{-1} and 8.0×10^6 M^{-1}, respectively. The binding of $(C1)_4$ is close to the K_i observed for the interaction of the native protein with the same substrate (2.5×10^6 M^{-1}) (not shown). Under these buffer conditions, the $(C2)_4$ tetramer binds poly[r(ϵA)] with the highest affinity ($K_i = 8.0 \times 10^6$ M^{-1}).

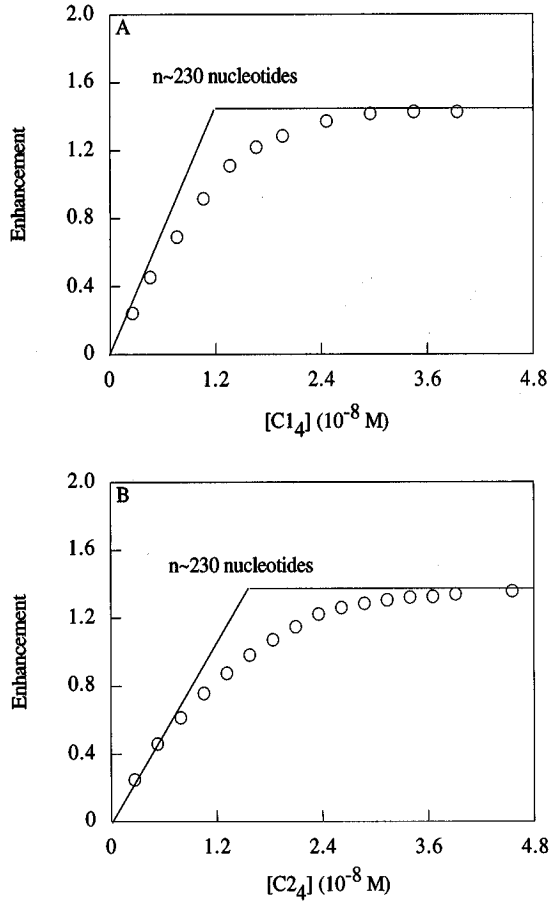

FIGURE 1: Stoichiometry of $(C1)_4$ and $(C2)_4$ binding to poly[r(ϵA)]. Titrations of 2.8 μM poly[r(ϵA)] with $(C1)_4$ (panel A) and of 2.7 μM poly[r(ϵA)] with $(C2)_4$ (panel B) were carried out in buffer B (10 mM Tris, pH 8.0, 1.0 mM EDTA, 100 mM NaCl). The binding site size (n) for each protein was estimated by extrapolation from the linear region of each curve (estimation of stoichiometric binding) to the point of apparent saturation indicated in the figure by the intersection of the two straight lines. The stoichiometry of binding for $(C1)_4$ and $(C2)_4$ at this point was approximately 230 nucleotides.

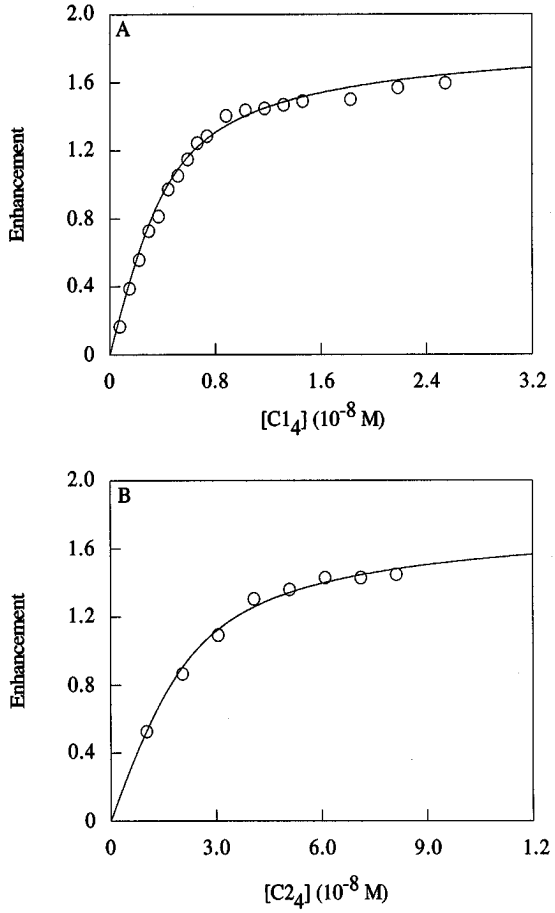

FIGURE 2: Binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly[r(ϵA)] in buffer B. Poly[r(ϵA)] (0.8 μM) was titrated with $(C1)_4$ (panel A) or $(C2)_4$ (panel B). Binding parameters were obtained by simulating theoretical curves (smooth line) to fit the experimental data (open circles) using the McGhee−von Hippel noncooperative model and assuming a site size of 230 nucleotides. The intrinsic association constants for the interaction of $(C1)_4$ and $(C2)_4$ with poly[r(ϵA)] were 3.9×10^6 and 8×10^6 M^{-1}, respectively.

The contiguous reiteration of 40S monoparticles along the length of nascent transcripts is consistent with a sequence-independent interaction of the core particle protein components with RNA. As pointed out in the introduction, it has been reported that C protein binds to homoribopolymers with widely different affinities, leading to the deduction that C protein preferentially binds *in vivo* to uridine rich sequences (Swanson & Dreyfuss, 1988a,b; Hamilton *et al.*, 1993; Olsen *et al.*, 1992; Wilusz & Shenk, 1990; Gorlach *et al.*, 1994). However, in those studies nonequilibrium binding assays were utilized that could not evaluate equilibrium binding parameters. To address this apparent paradox, we have used a series of competition binding experiments to identify the equilibrium binding parameters for the association of $(C1)_4$ and $(C2)_4$ to poly(G), poly(U), poly(C), and poly(A). The theoretical curves shown in Figure 2 were used to analyze competition titrations where the fluorescent probe substrate poly[r(ϵA)] was mixed with one of the competitor nucleic acids. To interpret the data from the competition binding assays, the two equilibria (binding to poly[r(ϵA)] or the competitor) were equated through the free protein concentration as shown in eq 1:

$$L = \nu_p\left\{K_p(1 - n_p\nu_p)\left[\frac{(2\omega_p - 1)(1 - n_p\nu_p) + \nu_p - R_p}{2(\omega_p - 1)(1 - n_p\nu_p)}\right]^{(n_p-1)}\left[\frac{1 - (n_p + 1)\nu_p + R_p}{2(1 - n_p\nu_p)}\right]^2\right\}^{-1}$$
$$= \nu_c\left\{K_c(1 - n_c\nu_c)\left[\frac{(2\omega_c - 1)(1 - n_c\nu_c) + \nu_c - R_c}{2(\omega_c - 1)(1 - n_c\nu_c)}\right]^{(n_c-1)}\left[\frac{1 - (n_c + 1)\nu_c + R_c}{2(1 - n_c\nu_c)}\right]^2\right\}^{-1} \quad (1)$$

where $R = [[1 - (n + 1)\nu]^2 + 4\omega\nu(1 - n\nu)]^{1/2}$, L is the free protein concentration, and K, n, ν, and ω are the association constant, site size, binding density, and cooperativity parameter for the interaction of the protein with poly[r(ϵA)] (subscript p) or competitor nucleic acid (subscript c), respectively. The strategy for simulating theoretical curves to fit experimental data from competition titrations involves finding the total protein concentration (L_T) required to produce the same fractional saturation (θ) of poly[r(ϵA)] in the presence of the competitor that was observed in the

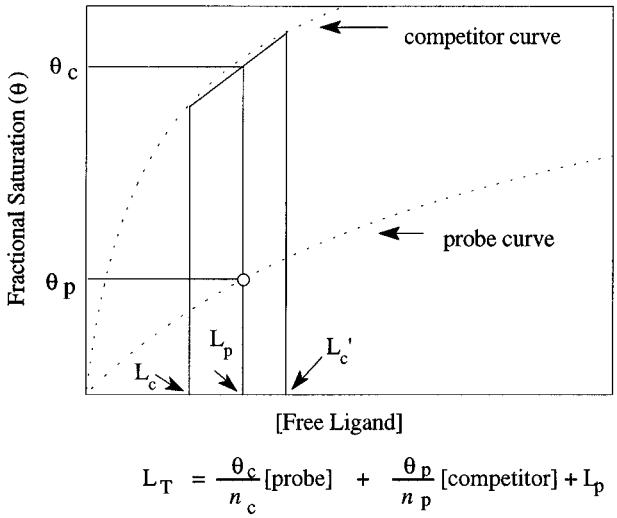

FIGURE 3: Method for determining binding constants from competition titrations (Kowalczykowski *et al.*, 1986). The binding of a hypothetical protein to poly[r(ϵA)] is calculated and plotted (probe curve). For a given set of binding parameters (K_i, ω, and n), a competitor curve is calculated and plotted. To determine the amount of total protein (L_T) required to produce a given fractional saturation of the probe (θ_p) in the presence of competitor, values of the free protein concentration on the competitor curve (L_c and $L_{c'}$) are found that bracket the free protein concentration on the probe curve (L_p). The fractional saturation of the competitor (θ_c) can then be determined by linear interpolation. The (L_T) for the competitor binding isotherm is then determined using the equation shown in the figure.

control titration in the absence of competitor. A program was developed for analyzing data from the competition titrations as outlined in Kowalczykowski *et al.* (1986). This analysis is presented graphically in Figure 3. Shown in Figure 3 is a theoretical curve corresponding to the binding of a hypothetical protein to poly[r(ϵA)] (probe curve). For any test value of K_i, n, or ω, a similar curve for a hypothetical competitor nucleic acid can be generated (competitor curve). Values of the free protein concentration on the competitor curve (L_c and $L_{c'}$) that bracket the free protein concentration on the probe curve (L_p) can then be determined. By linear interpolation, the fractional saturation of the competitor (θ_c) at L_p can be determined. The value of L_T required to produce the fractional saturation of the probe (θ_p) in the presence of the competitor is then calculated using eq 2:

$$L_T = \frac{\theta_c}{n_c}[\text{probe}] + \frac{\theta_p}{n_p}[\text{competitor}] + L_p \qquad (2)$$

where n_p and n_c are the binding site sizes of the protein to the probe and to the competitor nucleic acid, respectively (Kowalczykowski *et al.*, 1986). The goodness of fit for each set of binding parameters is evaluated by visual inspection and by minimizing the sum of the squared deviation for each data point between the theoretical and experimental values (Kowalczykowski *et al.*, 1986).

Figure 4 shows binding isotherms for the interaction of $(C1)_4$ and $(C2)_4$ tetramers with poly(G). The initial repression of the fluorescence enhancement is consistent with both proteins exhibiting a much higher affinity for binding poly(G) than poly[r(ϵA)]. The affinities of $(C1)_4$ and $(C2)_4$ to poly(G) are essentially the same, and approach stoichiometric binding (1.2×10^9 M^{-1} and 1.0×10^9 M^{-1}, respectively). The RNA:$(C2)_4$ ratio at probe saturation is much lower than expected if $(C2)_4$ binds poly(G) with the same affinity and

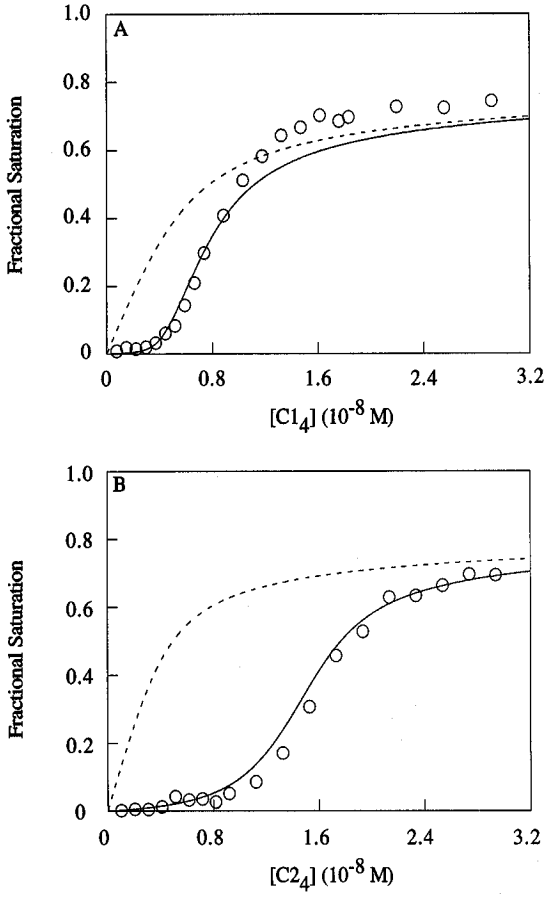

FIGURE 4: Determination of binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(G) in buffer B. Competition titrations between poly[r(ϵA)] and poly(G) were carried out to determine binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(G). Panel A shows data (open circles) from a titration of a mixture containing 1.4 μM poly[r(ϵA)] and 0.8 μM poly(G) with $(C1)_4$. The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel noncooperative model assuming a K_i of 1.2×10^9 M^{-1} and an n of 130 nucleotides. In panel B, a mixture of 0.8 μM poly[r(ϵA)] and 0.8 μM poly(G) was titrated with $(C2)_4$ (open circles). The theoretical curve through the data points (solid line) was also generated using the McGhee−von Hippel noncooperative model assuming a K_i of 1.0×10^9 M^{-1}, and an n of 42 nucleotides.

site size as the probe. This indicates that the site size for $(C2)_4$ binding to poly(G) is much smaller than that observed for the probe. This is quantitatively verified by the fit that defines a n of 42 nucleotides. In identical experiments, native C protein binds poly(G) with the same site size as $(C2)_4$ (not shown).

Analysis of competition titrations for the interaction of $(C1)_4$ and $(C2)_4$ tetramers with poly(U) is shown in Figure 5. The K_i for the interaction of $(C1)_4$ with poly(U) ($K = 1.2 \times 10^7$ M^{-1}) is almost identical to that observed for the protein's interaction with poly[r(ϵA)]. However, the binding site size for the interaction of $(C1)_4$ with poly(U) is 63 nucleotides, which defines an additional binding mode for the interaction of $(C1)_4$ with this homoribopolymer. The observed K_i for $(C1)_4$ binding poly(U) is approximately 100-fold less than that observed for the protein's interaction with poly(G). These observations contradict previous studies suggesting that among the four homoribopolymers, C protein binds poly(U) with the highest affinity (Swanson & Dreyfuss, 1988a). Like $(C1)_4$, $(C2)_4$ binds poly(U) and poly[r(ϵA)] with similar affinities and approximately 100-fold less than its affinity for poly(G).

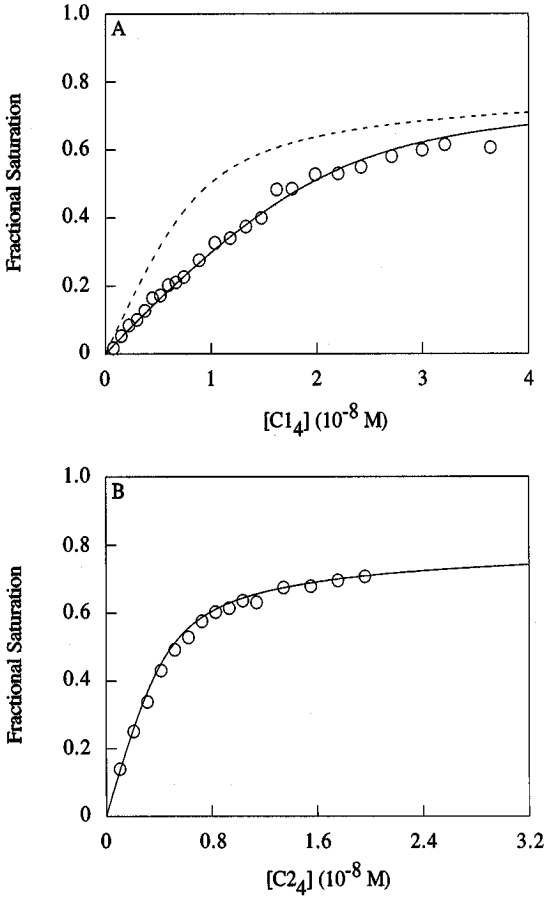

FIGURE 5: Determination of binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(U) in buffer B. Panel A shows data (open circles) from a titration of a mixture containing 1.4 μM poly[r(ϵA)] and 1.7 μM poly(U) with $(C1)_4$. The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel noncooperative model assuming a K_i of 1.2 $\times$ 10^7 M^{-1} and an n of 63 nucleotides. In panel B, a mixture of 0.8 μM poly[r(ϵA)] and 0.8 μM poly(U) was titrated with $(C2)_4$ (open circles). The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel noncooperative model assuming a K_i of 1 $\times$ 10^7 M^{-1} and an n of 230 nucleotides. The dashed line in panel A is the calculated curve that would be expected for the sum of the poly[r(ϵA)] and poly(U) concentrations if both proteins bound poly(U) with the same affinity and site size as poly[r(ϵA)]. In the case of $(C2)_4$, the dashed line superimposes the fit, indicating that this protein binds poly(U) in the same manner as poly[r(ϵA)].

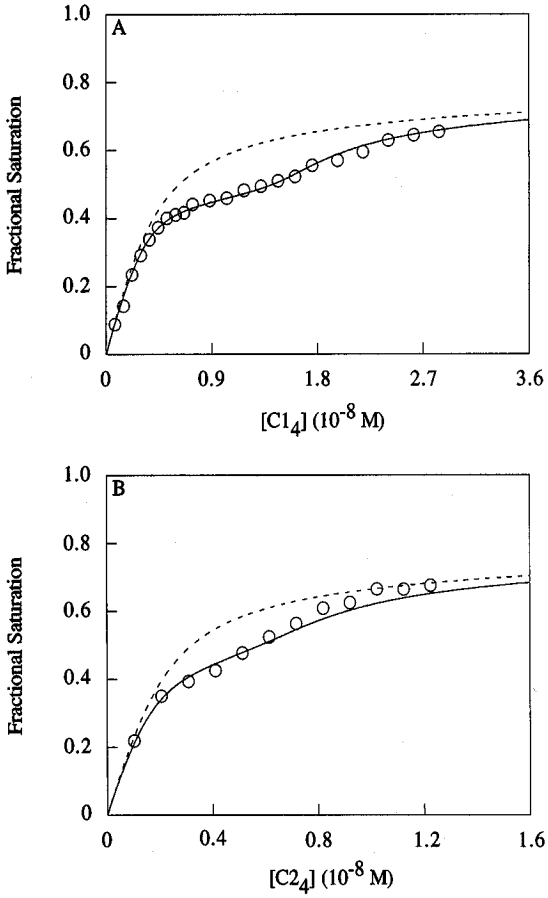

FIGURE 6: Determination of binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(A) in buffer B. Panel A shows data (open circles) from a titration of a mixture containing poly[r(ϵA)] (1.5 μM) and poly(A) (1.6 μM) with $(C1)_4$. The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel cooperative model assuming a K_i of 1.0 $\times$ 10^5 M^{-1}, a ω of 4400 and an n of 170 nucleotides. In panel B, a mixture of 0.8 μM poly[r(ϵA)] and 0.8 μM poly(U) was titrated with $(C2)_4$ (open circles). The theoretical curve through the data points (solid line) was also generated using the McGhee−von Hippel cooperative model assuming a K_i of 4.0 $\times$ 10^5 M^{-1}, a ω of 2000, and an n of 230 nucleotides. The dashed line in both panels represents the calculated curve for the interaction of both proteins with 1.5 μM (panel A) and 1.6 μM (panel B) of poly[r(ϵA)] in the absence of competitor.

However, the binding site size for the interaction of $(C2)_4$ with poly(U) is the same as that observed for poly[r(ϵA)] (Figure 5). While the occluded site sizes for the interaction of $(C1)_4$ and $(C2)_4$ with poly(U) are different, the magnitude of the association constants are essentially the same at 10^7 M^{-1}.

The C Proteins Bind RNA through a Highly Cooperative Binding Mode. The initial experiments suggesting that C protein binds poly(U) with high affinity also revealed a negligible affinity for poly(A) and poly(C) (Swanson & Dreyfuss, 1988a). Previous fluorescence binding experiments on an amino-terminal C protein fragment (containing the consensus RRM) indicated a lower affinity for poly(A) and poly(C) relative to poly(G) and poly(U) (Amrute *et al.*, 1994). Shown in Figure 6 are titrations with $(C1)_4$ and $(C2)_4$ where poly(A) is the competitor nucleic acid. The experimental data for the binding of both proteins to poly[r(ϵA)] in the presence of poly(A) indicate that both proteins exhibit multistate binding behavior with regard to the competitor nucleic acid. The multicomponent binding is characterized initially by a low-affinity binding mode (where the proteins preferentially bind poly[r(ϵA)]), followed by a high-affinity binding mode (where the competitor quickly becomes saturated through a narrow range of protein concentrations). These binding modes impart a biphasic character to the experimental data (Figure 6). The only multistate binding process where the low affinity component occurs first is cooperative binding. Thus, the low-affinity step reflects protein binding to noncontiguously bound sites in the region of low binding density followed by positive cooperative interactions at higher binding densities.

Previous studies have reported that native C protein and the amino-terminal fragment of C protein either do not bind to poly(C) or bind with very low affinity. The experimental findings presented in Figure 7 reveal that $(C1)_4$ and $(C2)_4$ bind poly(C) with very low intrinsic affinities but with very high cooperativity. The dashed lines in Figure 7A,B are the theoretical curves for the binding of $(C1)_4$ (panel A) and

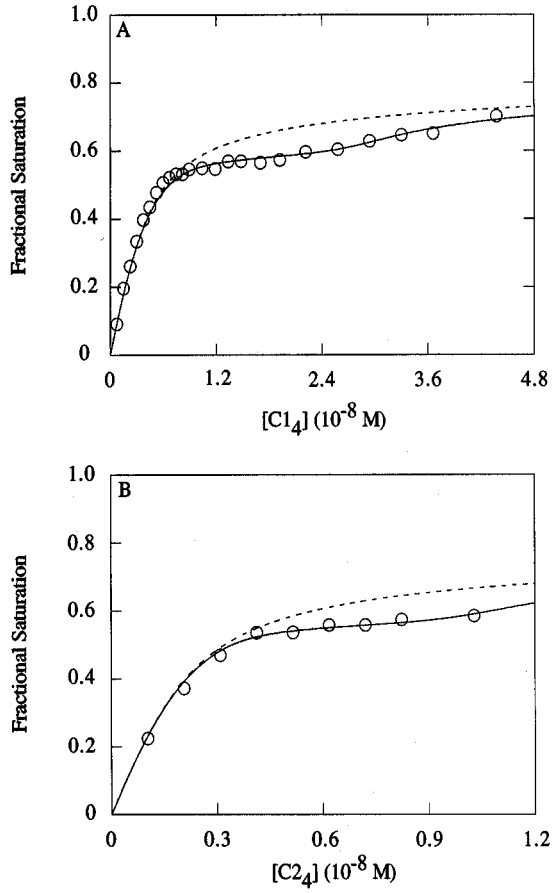

FIGURE 7: Determination of the binding parameters for the interaction of $(C1)_4$ and $(C2)_4$ with poly(C) in buffer B. Panel A shows data (open circles) from a titration of a mixture containing poly[r(ϵA)] (1.5 μM) and poly(C) (3.3 μM) with $(C1)_4$. The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel cooperative model assuming a $K\omega$ of 1.7×10^8 M^{-1} and an n of 200 nucleotides. In panel B, a mixture of 0.8 μM poly[r(ϵA)] and 0.8 μM poly(C) was titrated with $(C2)_4$ (open circles). The theoretical curve through the data points (solid line) was generated using the McGhee−von Hippel cooperative model assuming a $K\omega$ of 4.2×10^8 M^{-1} and an n of 150 nucleotides. The dashed line in both panels represents the calculated curve for the interaction of both proteins with 1.5 μM (panel A) or 0.8 μM (panel B) poly[r(ϵA)] in the absence of competitor.

$(C2)_4$ (panel B) to the probe substrate poly[r(ϵA)] in the absence of competitor (derived from fits of the experimental data in Figure 2). The binding of both proteins to poly[r(ϵA)] in the presence of poly(C) is qualitatively similar to the observations described in Figure 6 for the binding of both proteins to poly(A). Very low intrinsic affinities result in the partitioning of each protein onto poly[r(ϵA)] throughout a wide range of protein concentrations. The intrinsic affinities here are lower than those observed for the binding of each protein to poly(A). As a result, we can only define the maximum and minimum value for K_i and ω, respectively. For the interaction of $(C1)_4$ with poly(C), K_i is $\leq 1.5 \times 10^4$ M^{-1}, and the minimum value of ω is 11 000. $(C2)_4$ binds poly(A) with a $K_i \leq 5.0 \times 10^4$ M^{-1} and a minimum cooperativity of 8300. As a result of the high cooperativity observed for both proteins interaction with poly(C), saturation of this substrate occurs through a narrow range of protein concentrations once the threshold concentration is reached where the apparent affinity is $K\omega$. The binding data from these experiments are summarized in Table 1.

Table 1

	K_i (M^{-1})	ω	n	$K\omega$ (M^{-1})
		$(Cl)_4$		
poly(G)	1.2×10^9	na[a]	130	na
poly(U)	1.2×10^7	na	63	na
poly(C)	$\leq 1.5 \times 10^4$	$\geq 11\,000$	200	1.7×10^8
poly(A)	1.0×10^5	4400	170	4.4×10^8
		$(C2)_4$		
poly(G)	1.0×10^9	na	42	na
poly(U)	1.0×10^7	na	230	na
poly(C)	$\leq 5.0 \times 10^4$	≥ 8300	150	4.2×10^8
poly(A)	4.0×10^5	2000	230	8.0×10^8

[a] na = not applicable.

DISCUSSION

Proteins C1 and C2 are present in HeLa cells at a 3:1 molar ratio and exist as $(C1)_3C2$ heterotetramers (Barnett *et al.*, 1989; Huang *et al.*, 1994) that are restricted to the nucleus during interphase (Pinol-Roma & Dreyfuss, 1992). The failure to detect tetramers possessing more than one C2 polypeptide prompted the suggestion in a previous report that the 13 amino acid insert in C2 (at Gly 106) may in some way function in oligomerization to ensure heterotetramer stability (Barnett *et al.*, 1989). As shown here, however, the spontaneous oligomerization of C1 and C2 in bacterial cells to form highly stable homotetramers demonstrates that oligomerization is an intrinsic property of both proteins and that the 13 amino acid insert in C2 does not function to ensure tetramer stoichiometry. As shown here $(C1)_4$, $(C2)_4$, and $(C1)_3C2$ are indistinguishable based on their hydrodynamic and ultrastructural properties. They sediment at 5.6−5.8S in glycerol gradients with an apparent mass near 130 kDa, they elute from size-exclusion columns as asymmetric complexes with the same anomalously large Stokes radius, and they possess the same elution properties from strong anion-exchange columns. The recombinant and native proteins have essentially identical fluorescence excitation and emission spectra, characterized by a unique excited state tyrosine fluorescence emission (tyrosinate). All three proteins bind the fluorescent RNA substrate poly[r(ϵA)] with similar affinities and site sizes. These observations, along with quantitative comparative analysis of the binding properties of the recombinant and native proteins, are consistent with all three proteins being structurally as well as functionally equivalent.

The equilibrium binding studies described here demonstrate that $(C1)_4$ and $(C2)_4$ bind poly(A) and poly(C) with high cooperativity and with an occluded site size between 150 and 230 nucleotides. These findings are consistent with previous evidence indicating that, unless excluded by other factors, C protein has a clustered distribution on nascent transcripts (Conway *et al.*, 1988; Huang *et al.*, 1994; Rech *et al.*, 1995a). The cooperativity component of the overall binding affinity ($K\omega$) comprises the major component of the recombinant protein's affinity for single-stranded RNAs like poly(A) and poly(C). These RNAs are reported not to form highly stable secondary structures in solution as does poly(G) (discussed below). The equilibrium binding data reported here reveal that $(C1)_4$ and $(C2)_4$ bind poly(G) with an intrinsic affinity more than 100-fold higher than poly(U) and $\geq 10^4$-fold higher than poly(A) or poly(C). However, an inverse relationship exists between the magnitude of the

intrinsic association constant (K_i) and the cooperativity parameter (ω). For example, $(C1)_4$ binds poly(A) and poly(C) with relatively low intrinsic affinities (1.0×10^5 and $\geq 1.5 \times 10^4$ M^{-1}, respectively) but with very high cooperativity ($\omega = 4.4 \times 10^3$ and $\leq 11\,000$). Thus, the overall affinity (K_ω) for these substrates (4.4×10^8 and 1.7×10^8 M^{-1}) approximates their interaction with poly(G) (see Table 1). When the overall affinity is considered, the binding of $(C1)_4$ and $(C2)_4$ to poly(A) and poly(C) far exceeds the magnitude of their intrinsic interactions with poly(U). The characterization of C protein as a poly(U) binding protein was initially based on the finding that C protein−poly(U) complexes, and to a lesser extent C protein−poly(G) complexes, were found to resist dissociation at salt concentrations of 2 M NaCl (Swanson & Dreyfuss, 1988b). In this initial interpretation, affinity was strictly correlated with the ability of complexes to resist dissociation at high ionic strength (this, however, only measures the dependency of the association constant on ionic strength). In addition, because affinity is a measure of the magnitude of the equilibrium constant (including cooperativity), conclusions about this binding parameter could not be made from nonequilibrium experiments (i.e., experimental observations are taken when the system is not at steady state, and the concentrations of reactants and products are changing). Essentially these nonequilibrium experiments did not evaluate an equilibrium constant nor did they delineate or consider the affects of positive or negative cooperativity on binding. The studies to date suggesting that C protein is a poly(U) binding protein or that it specifically binds uridylate-rich sequences likely reflect measurements of the intrinsic affinity of C protein to isolated sites (see below).

High-affinity binding sites for recombinant C1 have been identified by selection/amplification from random pools of RNA sequences (SELEX experiments) (Gorlach *et al.*, 1994). In these experiments, the RNA substrates used were approximately 80 nucleotides in length and contained a randomized region of 20 nucleotides. The high-affinity consensus motif derived from this study consisted of five contiguous uridylates. In contrast, the findings described here demonstrate that $(C1)_4$ binds poly(G) with the highest intrinsic affinity among the four homoribopolymers. Similar results were reported previously for an amino-terminal fragment of C protein containing the consensus RNA binding domain (Amrute *et al.*, 1994). Our data also show that when cooperativity is considered, the overall affinity of $(C1)_4$ for poly(C) and poly(A) is 14 and 37 times higher than for poly(U) respectively. These findings bring into question the thermodynamic relevance of the SELEX-identified sequence for two fundamental reasons. First, the substrate RNAs were of inadequate length to support cooperative binding. Second, because the binding site size for the C protein tetramer exceeds the length of the SELEX substrates (20 nucleotides), an equilibrium was established where the protein functioned as the macromolecule (contains multiple sites for ligand binding) and the nucleic acid substrates were the ligand. Thus, the SELEX results may not reflect the physiologically relevant equilibrium since *in vivo* RNA is the macromolecule ranging in length mostly from 1000 to 20 000 nucleotides.

Under the conditions used in this study, poly(A) and poly(C) are reported to be mostly single-stranded molecules that belong to the A-RNA family possessing the C3′-endo nucleotide conformation (Saenger, 1988). On the other hand, poly(G) exists in a number of stable secondary structures including duplex, triplex, and quadruplex conformations (Saenger, 1988; Guschlbauer *et al.*, 1990). Poly(U) possesses a low degree of base stacking and also forms double-stranded hairpin structures that are stabilized by increasing ionic strength (Saenger, 1988). The absence of cooperative C protein binding to poly(G) and poly(U) and the smaller RNA binding site sizes observed for these substrates are likely to result from the fact that these RNAs form more stable alternative structures in solution in comparison to poly(A), poly(C), and most pre-mRNAs (Saenger, 1988; Guschlbauer *et al.*, 1990). Eukaryotic messenger RNA molecules approximate random polymers of the four nucleotides (Senapathy, 1986) and are rapidly hydrolyzed by single-strand-specific nucleases (Jelinek & Darnell, 1972; Calvet & Pederson, 1977; Kish & Pederson, 1978). The RNA binding site sizes that we observe for C protein to poly(A) and poly(C) range between 150 and 230 nucleotides and closely approximate the length of pre-mRNA bound per tetramer (235 nucleotides) based on previous *in vitro* reconstitution, hydrodynamic, and ultrastructural studies (Huang *et al.*, 1994; Rech *et al.*, 1995a). The high degree of cooperativity observed for the interaction of C protein with these two nucleic acid substrates is also consistant with experiments indicating that C protein is contiguously distributed along the length of nascent transcripts. These observations further suggest that the cooperative binding mode displayed by the interaction of $(C1)_4$ and $(C2)_4$ with poly(C) and poly(A) is the physiologically relevant interaction. Because the formation of hnRNP complexes on nascent transcripts is concomitant with transcription (suggesting that complex formation precedes and excludes the development of secondary structure in the RNA substrate), the poly(A) and poly(C) substrates (which lack secondary structure) more closely approximate pre-mRNA. It is likely that the core hnRNP proteins function in part to limit the degree of conformational flexibility in the pre-mRNA substrate.

In this study, we could not detect differences in the hydrodynamic properties of the $(C1)_4$ and $(C2)_4$ tetramers, and the binding data for most substrates are similar. $(C2)_4$, however, binds poly(G) with a significantly smaller site size than $(C1)_4$ (42 vs 133 nucleotides, respectively). Like $(C1)_4$ and $(C2)_4$, the native $(C1)_3C2$ tetramers bind poly(G) with the same high intrinsic affinity, yet like $(C2)_4$, the native protein also binds this substrate with a small site size (about 40 nucleotides). Thus, poly(G) has served as a probe substrate to demonstrate that $(C2)_4$ is not identical in every respect to $(C1)_4$, and that the native tetramer exhibits chimeric behavior in comparison to the two recombinant homotetramers.

Equilibrium binding studies conducted on an amino-terminal fragment of calf thymus C protein containing the conserved RNA binding domain [residues 9−102 (C RRM)] revealed that the single C RRM binds poly[r(ϵA)] with an association constant of 3.0×10^5 M^{-1} (at 100 mM NaCl) (Amrute *et al.*, 1994). In the studies described here, the $(C1)_4$ and $(C2)_4$ tetramers bind poly[r(ϵA)] with 27 and 13 times higher affinity than reported for the C RRM fragment using the same solution conditions. This indicates that multiple RRM's in the tetramer are involved in RNA binding. The differential affinity of the amino-terminal fragment for poly(G) over poly[r(ϵA)] was only 1.7 times higher. Here, the 100-fold higher affinity of the tetramer for poly(G) is

probably due to the synergistic effect of having four RNA binding domains. The amino-terminal fragment of C protein was reported to bind poly(U) with a slightly higher affinity than to the probe (1.3 times higher than poly[r(ϵA)]). In this study, $(C1)_4$ revealed a slight preference for poly(U) over poly[r(ϵA)] (2.6-fold higher). Unlike the noncooperative binding observed for the amino-terminal domain on all substrates, the C protein tetramer binds poly(A) and poly(C) in a highly cooperative mode. This indicates that the region of C protein involved in cooperative interactions resides in the carboxy-terminal two-thirds of the protein.

The RNA binding site size for an amino-terminal fragment of hnRNP A1 (1−184) containing two RNA binding domains is 14 ± 2 nucleotides (Shamoo *et al.*, 1994). The site size for the binding of the amino-terminal fragment of C protein described above (a single RRM) is approximately 7 ± 1 nucleotides (Amrute *et al.*, 1994). The small occluded site size observed with isolated RNA binding domains, compared with that of the C protein tetramer, suggests that lengths of RNA associated with the tetramer span some distance between RRM's and may not be in close contact with protein. In part, this may explain why C protein does not confer nuclease resistance to the bound substrate (Huang *et al.*, 1994). If all 4 RRM's function in RNA binding and the interactive site size for each RRM is 8 nucleotides, then approximately 50 nucleotides could span between the contact sites if the RRM's are distally and symmetrically oriented. Electron micrographs indicate that the tetramer is a four-lobed complex (Huang *et al.*, 1994; Rech *et al.*, 1995a), and crystallographic studies on the consensus RRM of snRNP-U1A demonstrate that the RRM's support the bound RNA on a peripheral surface (Poznanovic & Sevaljevic, 1986). The regions of RNA spanning between contact sites on the surface of the tetramer could render this RNA accessible for binding by other core proteins or by factors involved in RNA processing.

It is likely that in the absence of constraining factors, nascent transcripts would form a wide range of secondary structures in the nucleoplasm. RNA processing elements could therefore be incorporated into a number of stable inter- or intramolecular RNA complexes. An important function of the abundant hnRNP core particle proteins may therefore be to maintain nascent transcripts in a topologically single-stranded state. The binding data presented here indicate that the C protein tetramer binds approximately 220 nucleotides of pre-mRNA through a highly cooperative sequence-independent mechanism. These findings are consistent with their recovery from hnRNP complexes at saturation levels (1 tetramer per 220 nucleotides), and with the biochemical, ultrastructural, and immunochemical studies indicating that the hnRNP core proteins bind along the length of nascent transcripts.

ACKNOWLEDGMENT

We thank Kenneth R. Williams, Director of the W. M. Keck Foundation Biotechnology Resource Laboratory (Yale University), for his assistance with various analyses of native and recombinant C protein preparations. We also thank Norzehan Abdul-Manan for sending us a preparation of poly[r(ϵA)]. We thank Joe Beecham of the Department of Physiology and Biophysics at the Vanderbilt University College of Medicine for his helpful discussions. Finally, we thank Dr. James Patton of the Department of Molecular Biology, Vanderbilt University, for his valuable input into some of the recombinant experimental techniques.

REFERENCES

Ackers, G. K. (1964) *Biochemistry 3*, 723−730.
Amero, S. A., Raychaudhuri, G., Cass, C. L., van Venrooij, W. J., Habets, W. J., Krainer, A. R., & Beyer, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8409−8413.
Amrute, S. B., Abdul-Manan, Z., Pandey, V., Williams, K. R., & Modak, M. J. (1994) *Biochemistry 33*, 8282−8291.
Barnett, S. F., LeStourgeon, W. M., & Friedman, D. L. (1988) *J. Biochem. Biophys. Methods 16*, 87−98.
Barnett, S. F., Friedman, D. L., & LeStourgeon, W. M. (1989) *Mol. Cell Biol. 9*, 492−498.
Barnett, S. F., Theiry, T. A., & LeStourgeon, W. M. (1991) *Mol. Cell. Biol. 11*, 864−871.
Beyer, A. L., Christensen, M. E., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell 11*, 127−138.
Billings, P. B., & Martin, T. E. (1978) *Methods Cell Biol. 17*, 349−376
Burd, C. G., Swanson, M. S., Gorlach, M., & Dreyfuss G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9788−9792.
Calvet, J., & Pederson, T. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3705−3709.
Choi, Y. D., Grabowski, P. J., Sharp, P. A., & Dreyfuss, G. (1986) *Science 231*, 1534−1539.
Conway, G., Wooley, J., Bibring, T., & LeStourgeon, W. M. (1988) *Mol. Cell Biol. 8*, 2884−2895.
Dreyfuss, G. (1986) *Annu. Rev. Cell Biol. 2*, 457−495.
Faiferman, I., Hamilton, M. G., & Pogo, A. O. (1970) *Biochim. Biophys. Acta 204*, 550−563.
Gorlach, M., Burd, C. G., & Dreyfuss, G. (1994) *J. Biol. Chem. 269*, 23074−23078.
Guschlbauer, W., Chantot, J., & Thiele, D. (1990) *J. Biomol. Struct. Dyn. 8*, 491−511.
Hamilton, B. J., Nagy, E., Malter, J. S., Arrick, B. A., & Rigby, W. F. (1993) *J. Biol. Chem. 268*, 8881−8887.
Harris, S. G., Martin, T. E., & Smith, H. C. (1988a) *Mol. Cell. Biochem. 84*, 17−28.
Harris, S. G., Hoch, S. O., & Smith, H. C. (1988b) *Biochemistry 27*, 4595−4600.
Hendrick, D., & De Kloet, S. R. (1975) *Mol. Biol. Rep. 2*, 225−232.
Huang, M., & LeStourgeon, W. M. (1994) in *Cell Biology*: *A Laboratory Handbook* (Celis, J. E., Ed.) pp 641−656, Academic Press, San Diego, CA.
Huang, M., Rech, J. E., Northington, S. J., Mayeda, A., Krainer, A. R., & LeStourgeon, W. M. (1994) *Mol. Cell Biol. 14*, 518−533.
Jelinek, W., & Darnell, J. E. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2537−2541.
Kiseleva, E., Wurtz, T., Visa, N., & Daneholt, B. (1994) *EMBO J. 13*, 6052−6061.
Kish, V. M., & Pederson, T. (1978) *Methods Cell Biol. 16*, 377−411.
Kowalczykowski, S. C., Paul, L. S., Longerg, N., Newport, J. W., McSwiggen, J. A., & von Hippel, P. H. (1986) *Biochemistry 25*, 1226−1240.
LeStourgeon, W. M., Barnett, S. F., & Northington, S. J. (1990) in *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies* (Strauss, P. R., & Wilson, S. H., Eds) Vol. 2, pp 477−502, The Telford Press, Caldwell, NJ.
Lothstein, L., Arenstorf, H. P., Chung, S., Walker, B. W., Wooley, J. C., & LeStourgeon, W. M. (1985) *J. Cell Biol. 100*, 1570−1581.
Malcolm, D. B., & Sommerville, J. (1977) *J. Cell Sci. 24*, 143−165.
McGhee, J. D., & von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469−489.H
Merrill, B. M., Barnett, S. F., LeStourgeon, W. M., & Williams, K. R. (1989) *Nucleic Acids Res. 17*, 8441−8449.
Olsen, H. S., Cochrane, A. W., & Rosen, C. (1992) *Virology 191*, 709−715.

Pinol-Roma, S., & Dreyfuss, G. (1992) *Nature 355*, 730−732.
Poznanovic, G., & Sevaljevic, L. (1986) *Cell Biol. Int. Rep. 10*, 55−63.
Pullman, J. M., & Martin, T. E. (1983) *J. Cell Biol. 97*, 99−111.
Rech, J. E., Huang, M. H., LeStourgeon, W. M., & Flicker, P. F. (1995a) *J. Struct. Biol. 114*, 84−92.
Rech, J. E., LeStourgeon, W. M., & Flicker, P. F. (1995b) *J. Struct. Biol. 114*, 77−83.
Ross, A., J. B., Laws, W. R., Rousslang, K. W., & Wyssbrod, H. R. (1991) in *Topics in Fluorescence Spectroscopy* (Lakowics, J. R., Ed.) Vol. 3, pp 1−53, Plenum Press, New York.
Saenger, W. (1988) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Samarina, O., & Krichevskaya, A. (1981) *Cell Nucl. 9*, 1−48.
Sambrook, J., Fritch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Nolan, C., Ed.) Vol. 2, pp 14.15−14.21, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Senapathy, P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2133−2137.
Shamoo, Y., Abdul-Manan, N., Patten, A. M., Crawford, J. K., Pelligrini, M. C., & Williams K. R. (1994) *Biochemistry 33*, 8272−8281.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503−517.
Spann, P., Feinerman, M., Sperling, J., & Sperling, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 466−470.
Sperling, J., & Sperling, R. (1990) in *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies* (Strauss, P. R., & Wilson, S. H., Eds.) Vol. 2, pp 453−476, The Telford Press, Caldwell, N.J.
Steiner, R. F., Kinner, W., Lunasin, A., & Delac, J. (1973) *Biochim. Biophys. Acta 294*, 24−37.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Swanson, M. S., & Dreyfuss, G. (1988a) *Mol. Cell. Biol. 8*, 2237−2241.
Swanson, M. S., & Dreyfuss, G. (1988b) *EMBO J. 7*, 3519−3529.
Swanson, M. S., Nakagawa, T. Y., LeVan, K., & Dreyfuss G. (1987) *Mol. Cell. Biol. 7*, 1731−1739.
van Eekelen, C., Ohlsson, R., Philipson, L., Mariman, E., van Beek, R., & van Vanrooij, W. (1982) *Nucleic Acids Res. 10*, 7115−7131.
Vazquez-Nin, G. H., Echeverria, O. M., Martin, T. E., Luhrmann, R., & Fakan, S. (1994) *Eur. J. Cell Biol. 65*, 291−297.
Wilk, H. E., Angeli, G., & Schafer, K. P. (1983) *Biochemistry 22*, 4592−4600.
Wilusz, J., & Shenk, T. (1990) *Mol. Cell. Biol. 10*, 6397−6407.
Wooley, J., Chang, S. Y., Wall, J., & LeSourgeon, W. M. (1986) *Biophys. J. 49*, 17−19.

BI951974K