Universal Minicircle Sequence Binding Protein, a CCHC-type Zinc Finger Protein That Binds the Universal Minicircle Sequence of Trypanosomatids

Purification and Characterization*

Yehuda Tzfati, Hagai Abeliovich, Dana Avrhami, and Joseph Shlomai†

From the Department of Parasitology and Department of Cellular Biochemistry, The Hebrew University, Hadassah Medical School, Jerusalem 91010, Israel

Replication of kinetoplast DNA minicircles of trypanosomatid parasites initiates at a conserved 12-nucleotide sequence, termed the universal minicircle sequence (UMS, 5'-GGGGTTGTTGTA-3'). A single-stranded nucleic acid binding protein that binds specifically to this origin-associated sequence was purified to apparent homogeneity from Crithidia fasciculata cell extracts. This UMS-binding protein (UMSBP) is a dimer of 27.4 kDa with a 13.7-kDa protomer. UMSBP binds single-stranded DNA as well as single-stranded RNA but not double-stranded or four-stranded DNA structures. Stoichiometry analysis indicates the binding of UMSBP as a protein dimer to the UMS site. The five CCHC-type zinc finger motifs of UMSBP, predicted from its cDNA sequence, are similar to the CCHC motifs found in retroviral Gag polyproteins. The remarkable conservation of this motif in a family of proteins found in eukaryotic organisms from yeast and protozoa to mammals is discussed.

Kineto plast DNA (kDNA)† is a unique extrachromosomal DNA network found in the single mitochondria of flagellated protozoa of the family Trypanosomatidae. In Crithidia fasciculata, kDNA consists of about 5,000 DNA minicircles (2.5 kilobase pairs each) and about 50 DNA maxicircles (37 kilobase pairs each) interlocked topologically to form a huge DNA network (for review, see Refs. 1–3). Minicircles are heterogeneous in their nucleotide sequence but contain two short sequences, 70–100 base pairs apart, that are conserved in all the species studied so far: the dodecamer 5'-GGGGTTGTTGTA-3', and the hexamer sequence 5'-GGGGTTGTTGTA-3'. On the basis of in vivo observations, Enland and co-workers (4–8) have described the replication of kDNA minicircles as a process in which individual minicircles are detached from the central zone of the disc-shaped network, replicated, and reattached to the periphery of the disc. The network increases in size until it doubles and then divides and segregates into two daughter networks. Extensive studies of minicircle replication intermediates (9–16) have suggested that replication begins at the UMS site with synthesis of an RNA primer and proceeds by continuous elongation of the leading light strand (L-strand). A single gap of 6–10 nucleotides remains in the newly synthesized light strand at the UMS site (13) and is repaired only after replication of the minicircles and their reattachment to the network have been completed (8). Discontinuous synthesis of the lagging heavy strand (H-strand) starts when its origin, containing the conserved hexamer sequence, is exposed by the advancing replication fork. Highly gapped and nicked nascent H-strands are generated.

We have previously reported on the recognition of UMS by a unique sequence-specific single-stranded DNA binding protein from C. fasciculata (17) and on the isolation and analysis of the UMSBP-encoding cDNA (18). The amino acid sequence of the polypeptide, predicted from the cDNA, is 116 residues long and contains five Cys-X2-Cys-X4-His-X4-Cys (CCHC-type) zinc finger motifs. CCHC-type zinc finger motifs have been found in one or two copies in the retroviral nucleocapsid proteins and their Gag precursors (19) and in proteins of plant viruses and of eukaryotic cells. It has been suggested that this type of zinc finger is involved in binding of single-stranded nucleic acids (20). UMSBP belongs to a distinct group of cellular proteins that contains several (5–9) adjacent CCHC motifs, including cellular nucleic acid binding protein (CNBP) from human and mouse, which binds a G-rich single-stranded sequence of a sterol regulatory element (21); hexamer binding protein (HEXBp) from Leishmania major, which binds a G-rich single-stranded repeated sequence found in the 5'-untranslated region of the gene encoding GP63 (22); byr3 from Schizosaccharomyces pombe (23); and CnjB from Tetrahymena thermophila (24). The structure of the CCHC motifs of the HIV-1 Gag polyprotein was determined using NMR spectroscopy (25–28). These studies have revealed a very compact and well-defined structure, stabilized by coordination of the three cysteines and the histidine residue to the zinc ion and by extensive internal hydrogen bonding.

Here we describe the purification to apparent homogeneity of UMSBP from C. fasciculata cell extracts and the physical characteristics and specific nucleic acid binding properties of the protein. Finally we discuss the sequence and structure conservation of CCHC-type zinc finger motifs from UMSBP and other cellular proteins in reference to the structure and nucleic acid binding properties of the homologous retroviral motif.
EXPERIMENTAL PROCEDURES

Nucleic Acids, Nucleotides, Proteins, and Resins—Synthetic deoxy-
ligonucleotides were prepared by an Applied Biosystems oligo-
ucleotide synthesizer at the Bettlemann Laboratory of the Interdepart-
mental Division, Faculty of Medicine, the Hebrew University of J erusalem. Synthetic ribo-oligonucleotide (5’-GGGGUUUGGUAA-3’) was pre-
pared by New England Biolabs. Poly(dI-dC)poly(dI-dC) was purchased 
from Boehringer Mannheim. Phenyl-Sepharose was purchased from 
Sigma. Hydroxyapatite was from Bio-Rad and chromatofocusing resin 
(PBE™) and buffers (Polybuffer™) were from Pharmacia Biotech Inc. 
Radioactive nucleotides were purchased from DuPont NEN, polynuc-
otide kinase from New England Biolabs, and protein size markers from 
Sigma and Amersham Corp.

Cell Growth—Two hundred liters of C. fasciculata culture was grown 
at 29°C in a 250-liter industrial fermentor with 100 rpm stirring and 
air flow rate of 100 l/min. Cultures were harvested during logarithmic 
growth phase (0.5–1.0 

Nucleotide ligands (e.g., 4.5 g of yeast extract, 2 g of bovine serum albumin, 0.5 g of poly(dI-

dC)poly(dI-dC), and 0.2 ng of 32P-labeled 12-mer UMS H-strand

Gels were run on a polyacrylamide gel (16.5% T, 3% C separating gel; 10% T, 3% C 
spacers in TAE buffer (6.7 mM Tris acetate, 3.3 mM sodium 
acetate, 1 mM EDTA, pH 7.5). Electrophoresis was conducted at 4°C 
and 16 V/cm for 1.25 h. Gels were dried and exposed to x-ray films

Gel Filtration—UMSBP (Fraction VIIb) was filtered through a gel 
filtration G3000 SW HPLC column (0.75 × 60 cm; LKB) that was 
dipped in ice. The column was equilibrated and washed with 25 mM 
Tris-Cl, pH 7.5, 50 mM Na3SO4, 2 mM MgCl2, and 2 mM 
mercaptoethanol. Fractions of 200 µl were collected at a flow rate of 0.5 ml/min.

Glycerol Gradient Sedimentation—2.38 × 1011 units of UMSBP 
were centrifuged with 30 µl of Escherichia coli DNA polymerase I (5.6 S, 109 kDa), 150 µg of human hemoglobin (4.13 S, 
64.5 kDa), 10 units of horseradish peroxidase (3.85 S, 40 kDa), and 
150 µg of horse cytochrome C (2.1 S, 12 kDa). The 20-µl sample was 
layered onto a 5-ml 10–30% (v/v) linear glycerol gradient containing 
25 mM Tris-Cl, pH 7.5, 100 mM KCl, 2 mM MgCl2, and 1 mM DTT, 
and was centrifuged at 49,000 rpm for 39 h in a Kontes TLA-100 
50 rotor. Fractions of 0.3 ml were collected dropwise from the bottom of the tube and 
assayed for their UV absorbance at 260 nm. The fractions were layered onto a 5-ml 10–30% (v/v) linear glycerol gradient containing 
25 mM Tris-Cl, pH 7.5, 100 mM KCl, 2 mM MgCl2, and 1 mM DTT, and 
was centrifuged at 49,000 rpm for 39 h in a Kontes TLA-100 
50 rotor. Fractions of 0.3 ml were collected dropwise from the bottom of the tube and 
assayed for their UV absorbance at 260 nm. The fractions were

The Universal Minicircle Sequence Binding Protein

Purification

Purification of Crithidia fasciculata UMSBP—The electrophoretic mobility shift assay was used to monitor and quantify the specific binding of UMSBP to the UMS H-strand (17). C. fasciculata UMSBP was purified approximately 5,000-fold over the mixed cell lysate (Fraction I) to apparent homogeneity, with an overall yield of about 5% (Table I). The procedure yields 18.7 µg of pure UMSBP from 7 g of C. fasciculata wet cell paste (2 × 1013 cells).

The cleared cell lysate was fractionated by ammonium sul-
fate and then subjected to hydrophobic chromatography on phenyl-Sepharose and adsorption chromatography on hydroxy-
apatite. The following chromatofocusing step separated two forms of UMSBP (Table I, Fractions VIa and VIb) with estimated pl values of 7.25 and 6.25, respectively, and an apparent polypeptide mass of 12.6 and 13.7 kDa (under denaturing and reducing conditions (Fig. 1)). The partial sequencing of the shorter polypeptide chain by Edman degradation has revealed the absence of 11 amino acid residues at the protein N terminus compared to the sequence of the cDNA ORF (18). The identity of the longer polypeptide was verified by peptide mapping of the two polypeptide chains (not shown). Only the longer polypeptide was observed and co-chromatographed with UMS-

The protein concentrations were measured by the Coomassie blue reagent (38), and the purity of the preparations was determined by SDS-PAGE (39). The polypeptides were isolated from the gel using an electroelution apparatus (39) and analyzed by amino acid analysis (38) on a Beckman 120S amino acid analyzer.

Other methods—The 12-mer UMS H-strand was prepared as described 
previously (17). The DNA binding activity was measured by binding to the UMS H-strand in a cell lysate (Fraction I) that was centrifuged at 4°C for 39 h in a Kontes TLA-100 50 rotor. Fractions of 0.3 ml were collected dropwise from the bottom of the tube and assayed for their UV absorbance at 260 nm. The fractions were layered onto a 5-ml 10–30% (v/v) linear glycerol gradient containing 
25 mM Tris-Cl, pH 7.5, 100 mM KCl, 2 mM MgCl2, and 1 mM DTT, and 
was centrifuged at 49,000 rpm for 39 h in a Kontes TLA-100 
50 rotor. Fractions of 0.3 ml were collected dropwise from the bottom of the tube and 
assayed for their UV absorbance at 260 nm. The fractions were

Physical Properties—Molecular Weight and Subunit Structure—Some of the physical properties of C. fasciculata UMSBP are summarized in Table II. The purified protein migrates in SDS-polyacrylamide gels as a polypeptide band of 13.7 kDa (Fig. 1). This 13.7-kDa polypeptide chain co-fractionated with the specific UMS binding activity upon chromatography on phenyl-Sepharose (Fraction III) and hydroxyapatite (Fraction V), chromatofocusing (Fraction VI), G3000 SW HPLC gel filtration (Fig. 2) and glycerol gradient sedimentation (Fig. 3).
The Universal Minicircle Sequence Binding Protein

Purification was carried out following the specific binding of UMSBP to the UMS H-strand by the mobility-shift assay, as described under “Experimental Procedures.” Cell lysate (Fraction I, 1.56 g) was prepared from 7 g of C. fasciculata cell paste by gentle disruption of the Crithidia cell membrane, using a nonionic detergent in hypotonic solution, and was further fractionated by ammonium sulfate precipitation, as had been described previously (53). 525 mg of protein, precipitated with 40–60% (of saturation, at 0 °C) ammonium sulfate (Fraction II), were loaded onto a 35-ml phenyl-Sepharose column (1.35 × 18.5 cm) equilibrated with 50 mM Tris-Cl, pH 7.5, 1.5 M ammonium sulfate, 0.5 mM MgCl₂, and 2 mM β-mercaptoethanol. The column was washed with 2 bed volumes of the equilibration buffer, and proteins were eluted with a linear gradient of 1.5–0.7 M ammonium sulfate. UMSBP activity was recovered at 1.2–0.8 M ammonium sulfate to yield Fraction III (6.8 mg of protein). This fraction was then concentrated by loading it onto a second 3-ml phenyl-Sepharose column (0.9 × 4.7 cm), which was equilibrated and washed as above. Proteins were eluted with 50 mM Tris-Cl, pH 7.5, 0.5 mM MgCl₂, and 2 mM β-mercaptoethanol to yield Fraction IV. 3 mg of protein of Fraction IV were loaded onto a 4-ml hydroxyapatite column to yield Fraction V (154 mg of protein). 30.3 μg of Fraction V was adjusted to pH 8.0 with equilibration buffer (25 mM Tris acetate, pH 8.3, 2 mM MgCl₂, and 1 mM dithiothreitol), were loaded onto a pre-equilibrated 0.5-ml chromatofocusing (PBE™, Pharmacia) column (0.47 × 2.8 cm). Proteins were eluted with a 0.8–6.0 pH gradient, created within the column by the elution buffer (7.7% (v/v) Polybuffer 96™-acetate pH 6.0, 2 mM MgCl₂, and 1 mM dithiothreitol). UMS binding activity was recovered in two separated fractions: Fraction VIa at the approximate pH range of 7–7.5 (42.9 mg protein). Fraction VIa (10.6 mg) were loaded onto a 4-ml hydroxyapatite column (1.6 × 2.0 cm) equilibrated with 50 mM Tris-Cl, pH 7.5, 0.5 mM MgCl₂, and 2 mM β-mercaptoethanol. The column was washed with 2 bed volumes of the equilibration buffer, and proteins were eluted with a linear gradient of 0–75 mM potassium phosphate buffer, pH 7.5. UMS binding activity was recovered at 20–60 mM potassium phosphate, to yield Fraction VIb at the approximate pH range of 6.0–6.5 (6.0 mg of protein). 131 μg of protein of Fraction VIb, adjusted to pH 8.0 with equilibration buffer (25 mM Tris acetate, 0.5 M potassium phosphate, pH 7.5, 1.5 M ammonium sulfate, 0.5 mM MgCl₂, and 2 mM β-mercaptoethanol) was then concentrated by loading it onto a second 3-ml phenyl-Sepharose column (0.9 × 4.7 cm), which was equilibrated and washed as above. Proteins were eluted with 50 mM Tris-Cl, pH 7.5, 0.5 mM MgCl₂, and 2 mM β-mercaptoethanol to yield Fraction VIIa (10.6 mg of protein) and VIIb (4.8 mg of protein). Fraction VIIa and VIIb were loaded onto 50- and 100-ml hydroxyapatite columns, respectively, which were equilibrated with 25 mM Tris-Cl, pH 7.5, 2 mM MgCl₂, and 1 mM dithiothreitol. Protein fractions VIIa (10.6 mg) and VIIb (4.8 mg) were eluted with the equilibration buffer, containing 100 mM potassium phosphate, pH 7.5.

### Table I

| Fraction | Protein | Activity | Yield | Specific Activity | Purification |
|----------|---------|----------|-------|------------------|--------------|
| I. Clear lysate | 1.6 × 10⁶ | 49.9 | 0.03 |
| II. Ammonium sulfate | 5.3 × 10⁵ | 51.5 | (100) | 0.1 | 3 |
| III. Phenyl-Sepharose | 6.8 × 10⁵ | 43.0 | 85 | 6.5 | 202 |
| IV. Phenyl-Sepharose<sup>a</sup> | 36.1 | 72 | |
| V. Hydroxyapatite<sup>b</sup> | 322 | 17.2 | 33 | 53.3 | 1.7 × 10² |
| VIa. Chromatophocusing<sup>c</sup> | 147 | 10.1 | 20 | 69.0 | 2.2 × 10² |
| VIb. Chromatophocusing<sup>c</sup> | 19 | 2.1 | 4 | 110.0 | 3.4 × 10² |
| VIIa. Hydroxyapatite<sup>d</sup> | 40 | 3.6 | 7 | 90.5 | 2.8 × 10² |
| VIIb. Hydroxyapatite<sup>d</sup> | 16 | 2.6 | 5 | 158.5 | 5.0 × 10² |

<sup>a</sup> From this step onward, values were normalized for only part of the Fractions being used.

<sup>b</sup> UMSBP short polypeptide.

<sup>c</sup> UMSBP long polypeptide.

### Table II

| Physical properties of UMSBP |
|-----------------------------|
| Native mass | 27.4 kDa |
| Protomer mass | 13.7 kDa |
| s<sub>av</sub> | 2.4 S |
| Stokes Radius | 28 Å |
| ff<sub>f</sub> | 1.35 |
| Number of dimer UMSBP molecules | 12,000/Cell |

The apparent protomer mass, as measured by SDS-polyacrylamide gel electrophoresis, is similar to the one calculated on the basis of the UMSBP encoding cDNA (18). Gel filtration data yielded a Stokes radius of 28 Å, as calculated by the method of Siegel and Monty (34) (Fig. 2). A sedimentation coefficient of 2.37 S was measured in a 10–30% (v/v) glycerol gradient (Fig. 2). The Stokes radii were determined using bovine thyroglobulin (669 kDa). UMSBP was detected by the standard mobility-shift assay and protein markers by A<sub>280</sub>. The Stokes radius of UMSBP was interpolated from the linear plot of (−log K<sub>d</sub>)½ versus the known stokes radii values of the protein markers, as described by Siegel and Monty (34). Following the method of Martin and Ames (35), the apparent native mass of the protein, calculated from the experimental sedimentation coefficient and Stokes radius (assuming a partial specific volume of 0.725 ml/g) (34), is estimated as 27.4 kDa. The frictional coefficient (ff<sub>f</sub>), calculated by the method of Siegel and Monty (34), is 1.35, indicating an axial ratio of approximately 1:7.

These data suggest that the native C. fasciculata UMSBP is a homodimer with a protomer mass of 13.7 kDa. On the basis of...
Protein markers and their sedimentation coefficients (gradient and assayed as described under "Experimental Procedures.") Protein markers and their sedimentation coefficients (s_{20,w}) were E. coli DNA polymerase 1 (6.6 S), human hemoglobin (4.13 S), horseradish peroxidase (3.85 S), and horse cytochrome C (2.1 S). The sedimentation coefficient was interpolated from the linear plot of the s_{20,w} values of the markers, as described by Martin and Ames (35).

the protein activity in cell extracts, the apparent molecular weight, and the specific activity of the pure protein, we estimate the presence of approximately 12,000 UMSBP molecules/Crithidia cell.

Characteristics of the Binding Reaction—Generation of UMSBP-DNA complexes, as monitored by the mobility shift assay, is greatly enhanced in the presence of bovine serum albumin at a concentration of 1 mg/ml. Chelation of divalent cations by EDTA prevents the formation of protein-DNA complexes, indicating the possible involvement of metal ions in the putative zinc fingers. Optimal binding activity is observed at 60 mM KCl. Higher concentrations of monovalent ions are inhibitory, with 50% inhibition observed in the presence of 200 mM KCl or 150 mM NaCl. Glycerol concentrations up to 50% (v/v) enhance complex stability upon electrophoresis in native polyacrylamide gels. Optimal electrophoresis temperature is at the range of 2–5°C. At temperatures higher than 8°C, dissociation of protein-DNA complexes during electrophoresis is fast and results in smeared bands. Raising the polyacrylamide gel concentration from 4 to 8% stabilizes protein-DNA complexes and reduces their dissociation, presumably due to caging effect (36, 37).

Structural Features of the DNA Ligand—We have previously shown (17) the specific binding of UMSBP to a G-rich single-stranded sequence. The equilibrium binding constant measured for the interaction of UMSBP with a 12-mer oligonucleotide comprising the H-strand sequence of UMS, is 2.5 x 10^{9} M^{-1} (Fig. 4). A similar (2.6 x 10^{9} M^{-1}) equilibrium binding constant was measured with a 40-mer oligonucleotide that consists of the H-strand of UMS and its flanking sequences at the origin region of the C. fasciculata kDNA minicircle (Fig. 4). This observation indicates that neighboring sequences flanking UMS at the minicircle origin site, in their single-stranded conformation, have no significant effect on the binding of UMSBP.

G-rich sequences similar to UMS, such as those of eukaryotic telomere termini, retroviral RNA genome dimerization site, gene regulatory elements, and immunoglobulins switch regions, form in vitro special four-stranded (quadruplex) DNA structures. These structures, known as G-quartets or G4-DNA, are stabilized by Hoogsteen base pairing (38–42). Several proteins that have been discovered recently, bind specifically these special conformations (43, 44). Considering the specific binding of UMSBP to a G-rich ligand that may potentially form a four-stranded structure, we have explored the possibility that such a conformation is recognized by the protein. Since we could not detect stable quadruplexes formed in vitro by the 12-mer UMS H-strand oligonucleotide, we have used for this purpose a similar oligonucleotide containing the repeated Tetrahymena telomeric sequence 5'-GGGGTTGGGGTT-3'. UMSBP binds tightly to this telomeric sequence (Fig. 4 and Ref. 17). This oligonucleotide adopts two different DNA conformations that migrate as two different bands upon electrophoresis in a native polyacrylamide gel. The lower mobility band corresponds to the quadruplex structure, which is composed of two oligonucleotide molecules in a fold-back conformation (38, 45), while the higher mobility band represents the monomeric structure. Mobility-shift analyses (Fig. 5) clearly demonstrate that UMSBP binds only the higher mobility monomeric molecules, but not the lower mobility four-stranded dimers.

Stoichiometry of UMS DNA and UMSBP in the Protein-DNA Complex—On the basis of its encoding cDNA (18), UMSBP contains five CCHC-type zinc finger motifs. Thus, the native UMSBP homodimer isolated from C. fasciculata may contain a total of 10 such zinc finger structures. The functional role of each of these potential zinc fingers is yet to be studied. However, understanding the specific UMSBP-UMS interactions requires an accurate estimation of the stoichiometry of the protein and DNA reactants in this specific nucleoprotein complex.

No binding of a dimeric DNA ligand (such as G4-DNA) by UMSBP could be observed (Fig. 5). However, we have further explored the possibility that a single UMSBP molecule may bind simultaneously more than one UMS site. To address this question, we have used two DNA ligands that contain the 12-mer UMS sequence but differ in their length. The oligonucleotide UMS-H12 contains only the 12-mer H-strand of UMS,
while the 40-mer UMS-H40 contains the UMS 12-mer and its flanking sequence at the minicircle H-strand. Whereas both DNA ligands are tightly bound by UMSBP (equilibrium binding constants measured for the two protein-DNA interactions were almost identical (Fig. 4)), the two protein-DNA complexes differ in their electrophoretic mobility in native polyacrylamide gels. If UMSBP binds only one UMS site, then two types of protein-DNA complexes could be expected: UMSBP-(UMS-H12) and UMSBP-(UMS-H40). However, if the complex contains two UMS elements, then three types of complexes may be expected: UMSBP-(UMS-H12)₂, UMSBP-(UMS-H40)₂ as well as UMSBP-(UMS-H12)₁(UMS-H40)₁. Fig. 6 describes the results of such an experiment in which the oligonucleotides UMS-H12 and UMS-H40 were mixed together at various molar ratios as indicated, heat denatured in order to disrupt any pre-existing higher order structures, and used as radioactive probes in an electrophoretic mobility shift experiment with UMSBP. Reciprocal titration of one species of UMS-DNA over the other at the various molar ratios, yields only two types of protein-DNA complexes. No additional species of protein-DNA complexes could be detected, indicating that only one DNA molecule is present in the UMSBP-UMS complex.

To determine the precise number of UMSBP monomers that bind a single UMS element in the complex, we have conducted a mobility-shift electrophoresis analysis of the protein-DNA complexes using an ⁳⁵S-labeled UMSBP and ³²P-⁵'-labeled UMS DNA. We have measured a value of 2.1 UMSBP monomer/UMS site (Fig. 7), indicating the apparent binding of two UMSBP monomers to each UMS binding site and suggesting that UMSBP binds to DNA as a protein dimer.

Conservation of the CCHC-Type Zinc Finger Motif—The amino acid sequence of UMSBP, as predicted from the cDNA (18), contains five CCHC-type zinc finger motifs. The sequence conservation of this motif in a growing group of cellular proteins including UMSBP is demonstrated in Fig. 8. The 35 CCHC motifs of UMSBP and HEXBP (from flagellated protozoa), CnjB (from ciliated protozoa), byr3 (from yeast), and CNBP (from mammals) were found to be remarkably conserved. Apart from the conservation of the cysteines and histidine, which apparently function in the coordination of a zinc base, glycine residues are conserved at positions 5 and 8 of the motif. The presence of these glycine residues may reflect a requirement for small, sterically nondemanding residues at these positions of the compact motif structure, as suggested for the retroviral motifs (20). The conservation of a proline residue at position 15 may reflect a structural conservation of a turn in the backbone of the motif. An aromatic residue (tyrosine or phenylalanine) is conserved at position 2, a hydrophobic residue at position 10, and serine or alanine at position 11. The same residues at the same positions, were found by South and Summers (25) to form a hydrophobic cleft in the HIV-1 Gag motif within which the DNA ligand binds. The aromatic and alanine residues, at the same conserved positions (2 and 11) in the retroviral motif, form specific hydrogen bonds with a guanine base of the DNA ligand. However, another basic residue (lysine) that is present immediately before the first cysteine of the Gag motif and forms a specific contact with the guanine base is not conserved in the CCHC motifs of these cellular proteins. Instead, a basic residue that might serve the same function is conserved at position 3 of their motifs. Another basic
Residue is conserved at position 12. The side chain of an arginine residue at this position of the Gag motif was found to form a nonspecific electrostatic interaction with the phosphodiester backbone of the ligand. Finally, an acid residue (aspartate or glutamate) is conserved at position 13.

Overall, we have found a high degree of conservation in 13 out of the 15 positions of the CCHC motifs of this family of eukaryotic cellular proteins. This remarkable conservation can be explained in light of the functions found by South and Summers for the same residues at the same positions in the HIV-1 Gag motif (25).

Binding of UMSBP to Single-stranded RNA—It has been shown that in addition to its RNA ligand, the retroviral CCHC-type zinc finger can bind a single-stranded DNA analog (25). Considering the remarkable homology between the CCHC motifs of the retroviral Gag and those of UMSBP, it was expected that UMSBP would bind the single-stranded RNA ligand. To explore this possibility, we have used the 12-mer ribo-oligonucleotide analog of the G-rich strand of UMS (5'-GGGGUUUGGUGUA-3') in a mobility-shift assay. Fig. 9 demonstrates that UMSBP indeed binds the RNA analog of the UMS H-strand. Binding competition experiments, using an RNA competitor of a higher sequence complexity than that of UMS (Fig. 10), clearly demonstrates the sequence specificity of UMSBP interaction with the single-stranded RNA ligand. The equilibrium binding constant measured for this protein–RNA interaction (K = 2.0 \times 10^{10} \text{ M}^{-1}) is higher than the value measured with the DNA ligand (Fig. 4). Inasmuch as the CCHC-type zinc finger motif is conserved in the group of cellular proteins from yeast and protozoa to mammals, the presumption is strong that other proteins in this group also bind both types of single-stranded nucleic acids. Whether this RNA binding activity of UMSBP, which is apparently an intrinsic property of the CCHC-type zinc finger motifs, has any physiological significance, is yet to be determined.

DISCUSSION

During the S-phase of the trypanosomatid cell cycle, two highly interlocked kDNA catenanes, one composed of minicircles and the other of maxicircles, replicate at the same
time and at the same location. Thus, replication and assembly of the two types of these topologically linked kDNA circles requires a strict coordination between their replication mechanisms. Recently, two copies of an 11-mer sequence identical to UMS (apart from its 3′-terminal residue) were found in the maxicircle variable region of Trypanosoma brucei (46, 47). The presence of this origin-associated sequence in both minicircles and maxicircles may provide a clue for understanding this coordination at the replication initiation step. A specific origin-binding protein that interacts with the origin-associated UMS, is a likely candidate to function in the process of replication initiation and may play a role in a mechanism that coordinates kDNA minicircle and maxicircle replication. It is within this context that we had searched for and isolated a UMS-binding protein from C. fasciculata cell extracts. Since the 3′-terminal residue of UMS is insignificant for specific binding by UMSBP (17), we expect that both the 12-mer UMS of the minicircles and the homologous 11-mer sequence of the maxicircles would be equally bound by the protein. The conservation of UMSBP binding sites in both maxicircles and minicircles supports a possible role for UMSBP in coordinating the replication of the two types of circles. Since UMS resides within a duplex DNA molecule, binding of UMSBP requires the melting of this sequence. We have recently found that UMSBP binds to native DNA minicircles and that the origin-associated UMS element resides within an unwound or otherwise sharply distorted DNA structure. We have shown here that UMSBP can bind a UMS RNA analog, as implied by the remarkable homology of the CCHC motifs from UMSBP and the retroviral Gag polyproteins. Whether UMS is indeed transcribed in the trypanosomatid cell and a UMS RNA ligand is actually available for binding by UMSBP, is yet unknown. Further investigation is required to determine the in vivo binding target and the biological function of UMSBP.

G-rich sequences similar to the UMS, such as those of telomeres (38, 39), HIV-1 RNA genome dimerization site (40, 41), IgG switch region (42, 43), and others (44, 49, 50) form in vitro special four-stranded structures known as quadruplexes or G-quartets. Several DNA-binding proteins were recently found to interact specifically with a G-quartet structure (43, 44, 51, 52). Although UMSBP binds exclusively to single-stranded nucleic acid and may play a role in a mechanism that coordinates kDNA minicircle and maxicircle replication. It is within this context that we had searched for and isolated a UMS-binding protein from C. fasciculata cell extracts. Since the 3′-terminal residue of UMS is insignificant for specific binding by UMSBP (17), we expect that both the 12-mer UMS of the minicircles and the homologous 11-mer sequence of the maxicircles would be equally bound by the protein. The conservation of UMSBP binding sites in both maxicircles and minicircles supports a possible role for UMSBP in coordinating the replication of the two types of circles. Since UMS resides within a duplex DNA molecule, binding of UMSBP requires the melting of this sequence. We have recently found that UMSBP binds to native DNA minicircles and that the origin-associated UMS element resides within an unwound or otherwise sharply distorted DNA structure. We have shown here that UMSBP can bind a UMS RNA analog, as implied by the remarkable homology of the CCHC motifs from UMSBP and the retroviral Gag polyproteins. Whether UMS is indeed transcribed in the trypanosomatid cell and a UMS RNA ligand is actually available for binding by UMSBP, is yet unknown. Further investigation is required to determine the in vivo binding target and the biological function of UMSBP.

G-rich sequences similar to the UMS, such as those of telomeres (38, 39), HIV-1 RNA genome dimerization site (40, 41), IgG switch region (42, 43), and others (44, 49, 50) form in vitro special four-stranded structures known as quadruplexes or G-quartets. Several DNA-binding proteins were recently found to interact specifically with a G-quartet structure (43, 44, 51, 52). Although UMSBP binds exclusively to single-stranded nucleic acid and may play a role in a mechanism that coordinates kDNA minicircle and maxicircle replication. It is within this context that we had searched for and isolated a UMS-binding protein from C. fasciculata cell extracts. Since the 3′-terminal residue of UMS is insignificant for specific binding by UMSBP (17), we expect that both the 12-mer UMS of the minicircles and the homologous 11-mer sequence of the maxicircles would be equally bound by the protein. The conservation of UMSBP binding sites in both maxicircles and minicircles supports a possible role for UMSBP in coordinating the replication of the two types of circles. Since UMS resides within a duplex DNA molecule, binding of UMSBP requires the melting of this sequence. We have recently found that UMSBP binds to native DNA minicircles and that the origin-associated UMS element resides within an unwound or otherwise sharply distorted DNA structure. We have shown here that UMSBP can bind a UMS RNA analog, as implied by the remarkable homology of the CCHC motifs from UMSBP and the retroviral Gag polyproteins. Whether UMS is indeed transcribed in the trypanosomatid cell and a UMS RNA ligand is actually available for binding by UMSBP, is yet unknown. Further investigation is required to determine the in vivo binding target and the biological function of UMSBP.