Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10450–10454, December 1991
Biochemistry

A triplex DNA-binding protein from human cells: Purification and characterization

[three-stranded DNA/DNA–protein interactions/(A+T)-rich DNA/TAT triplex/triplex DNA affinity chromatography]

RYOITI KIYAMA AND R. DANIEL CAMERINI-OTERO

Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 10, Room 9D-15, Bethesda, MD 20892

Communicated by Gary Felsenfeld, July 30, 1991 (received for review March 18, 1991)

ABSTRACT **A protein that binds to an oligonucleotide triplex, $(dT)_{34}$·$(dA)_{34}$·$(dT)_{34}$ (TAT triplex), was purified from HeLa cells by a combination of conventional column chromatography and triplex DNA affinity chromatography. The protein has an apparent molecular mass of 55 kDa on sodium dodecyl sulfate/polyacrylamide gels. Although the protein has an affinity for AT duplex and TAT triplex, a higher affinity for TAT triplex was demonstrated by comparing the elution profiles from an AT duplex and a TAT triplex affinity column. The protein has a moderate affinity for poly(dA-dG)·poly(dT-dC) and a very low affinity for single-strand $(dA)_{34}$ or $(dT)_{34}$ and various sources of duplex DNA including poly(dA-dT)·poly(dA-dT). The possible biological function of this triplex DNA-binding protein is discussed.**

Nucleic acid triplexes were first described over three decades ago by Felsenfeld *et al.* (1). That structure was a three-stranded polyribonucleotide in which one strand was polyadenylic acid and the two other strands were polyuridylic acid. Several years later Arnott and Selsing (2), using x-ray fiber diffraction, concluded that, as in the RNA triplex, in the analogous DNA triplex [poly(dT)·poly(dA)·poly(dT)] the major structural features were that one of the polypyrimidine strands is wrapped around the duplex in the major groove. The base pairing pattern of the base triplets is as shown in Fig. 1*A*. Additional studies using nuclear magnetic resonance (3) and chemical probes (4) have confirmed and extended these earlier studies and, in particular, have confirmed the Hoogsteen base pairing scheme involved. In the meantime, several other intermolecular (5, 6) and intramolecular (H form; refs. 7, 8) triplexes between homopurine·homopyrimidine duplexes and a third strand were described. In all such triplexes the third strand is antiparallel to the identical strand in the duplex. Recently, we have described an intermolecular triplex that is probably an intermediate in homologous recombination and in which the third strand is parallel to the identical strand in the duplex (9).

Recently there has been renewed interest in those triplexes containing homopurine·homopyrimidine duplexes because such sequences are found in or close to transcriptional units and recombination hot spots in eukaryotes (see ref. 10 for review). Furthermore, these sequences are often responsible for the presence of S1-hypersensitive sites in some of the 5′ flanking regions of eukaryotic genes (11, 12). Also, relatively long stretches ($n \geq 10$) of homopurine·homopyrimidine sequences are known to be 3.5–4 times overrepresented in higher eukaryotes and eukaryotic viruses (13, 14). Among the homopurine·homopyrimidine sequences, (A+T)-rich tracts have been also implicated as anchorage sites for chromatin structure (15, 16).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

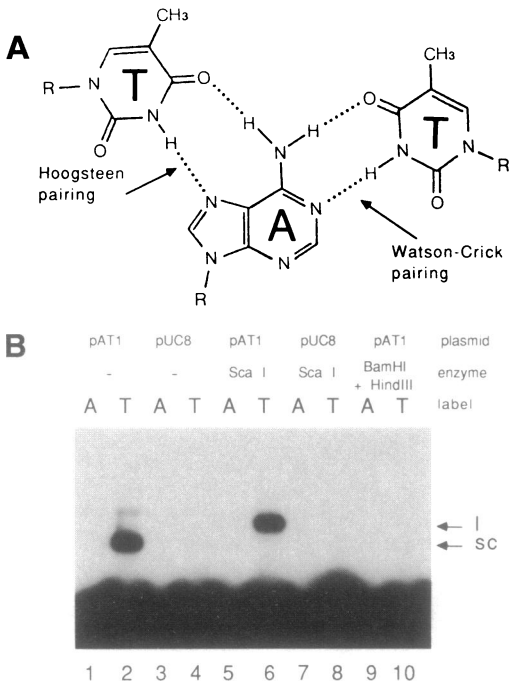

FIG. 1. Structure and formation of a triplex DNA. (*A*) Chemical structure of the TAT triplet bases. (*B*) Formation of triplex DNA. Plasmid pAT1, containing a $(dA)_{34}$·$(dT)_{34}$ tract (13.5 nM), before (lanes 1–4) or after (lanes 5–10) the indicated restriction enzyme digestion, was incubated with 100 nM ^{32}P-labeled $(dA)_{34}$ (lanes 1, 3, 5, 7, and 9) or 100 nM ^{32}P-labeled $(dT)_{34}$ (lanes 2, 4, 6, 8, and 10) at 37°C for 30 min and electrophoresed on a 0.7% agarose gel in TBM buffer at 4°C overnight. After staining with ethidium bromide, the gel was fixed with 10% (wt/vol) trichloroacetic acid, dried, and autoradiographed. Positions of linear (l) and supercoiled (sc) DNA are shown.

Although there is only circumstantial evidence (17) of a role for homopurine·homopyrimidine triplexes in cells, some recent biochemical data suggest possible biological functions. Hogan and colleagues (18) showed a correlation between triplex formation and repression of c-*myc* transcription *in vitro*. In a similar vein, micromolar concentrations of pyrimidine oligodeoxyribonucleotides were shown to block recognition of duplex DNA by various proteins, including a eukaryotic transcription factor (19).

Although no proteins have been described that preferentially bind to triplex DNA, some proteins are known to bind to homopurine·homopyrimidine sequences. In addition to histones and topoisomerase II, other nuclear proteins associated with chromosomal organization such as simian α-protein (20), *Drosophila* D1 protein (21), *Dictyostelium* proteins BP1 and BP2 (22), and yeast datin (23) also bind preferentially to oligo(dA)·oligo(dT) sequences.

To elucidate the function of triplex DNA, we have asked whether proteins can be found in the nucleus of cells that bind preferentially to DNA triplexes. We report here the purification and characterization of such a human nuclear protein.

MATERIALS AND METHODS

Materials. Synthetic duplex polynucleotides, *Escherichia coli* DNA, and CNBr-activated Sepharose 4B were purchased from Pharmacia. Oligonucleotides $(dA)_{34}$ and $(dT)_{34}$ were synthesized on an Applied Biosystems 380B DNA synthesizer. A pBR322 sequence (24), nucleotides 61–94 (34-mer, 50% purines) and its complementary sequence, were selected as a control DNA. Plasmid pAT1 was constructed from pUC8 by cloning a $(dA)_{34}$·$(dT)_{34}$ sequence flanked by *Bam*HI and *Hin*dIII ends into the polylinker region.

Preparation of Nuclear Extracts. Nuclear extracts were made from 2.5×10^{10} of HeLa cells according to Dignam *et al.* (25) with minor modifications. Briefly, cells were washed three times with phosphate-buffered saline (PBS), washed once with 600 ml of buffer A (10 mM Hepes, pH 7.5/1.5 mM $MgCl_2$/10 mM KCl/0.5 mM dithiothreitol), and resuspended in 100 ml of buffer A. After homogenization and centrifugation (2000 rpm, 10 min), nuclei were washed once with 200 ml of buffer A and resuspended in 200 ml of buffer C [20 mM Hepes, pH 7.5/20% (vol/vol) glycerol/0.42 M NaCl/1.5 mM $MgCl_2$/0.2 mM EDTA/0.5 mM phenylmethylsulfonyl fluoride/0.5 mM dithiothreitol]. After 30 min of constant stirring, the soluble fraction (S-100) was recovered from the suspension by two consecutive centrifugations: 8000 rpm for 10 min (Sorvall HB-4) and 36,000 rpm for 1 hr (Beckman 70Ti). All extraction procedures were carried out at 4°C.

Purification of Triplex DNA-Binding Protein. The S-100 fraction was diluted with 2 vol of buffer B [15 mM Hepes, pH 7.5/10% (wt/vol) glycerol/3 mM $MgCl_2$/0.5 mM EDTA/0.1 mM EGTA/0.2 mM phenylmethylsulfonyl fluoride/1 mM 2-mercaptoethanol] and applied to S Sepharose (Pharmacia; bed vol, 50 ml) for a batch extraction. Proteins were eluted sequentially with 100 ml of 150 mM, 500 mM, and 1 M NaCl in buffer B (four fractions each). The 500 mM fraction was fractionated and concentrated by stepwise precipitations with 40–70% saturated ammonium sulfate and loaded onto a hydroxylapatite column (bed vol, 6 ml) equilibrated with 50 mM phosphate buffer, pH 6.8/10% (wt/vol) glycerol/0.5 mM EDTA/0.1 mM EGTA/0.2 mM phenylmethylsulfonyl fluoride/1 mM 2-mercaptoethanol. Proteins were eluted with a total of 12 ml of 200 mM, 300 mM, 400 mM, and 500 mM phosphate buffer (four fractions each). The 400 mM fraction was dialyzed against 0.15 M NaCl plus buffer B and loaded onto a triplex DNA Sepharose column (bed vol, 1 ml) equilibrated with the same buffer. Proteins were eluted stepwise with a total of 2 ml of 0.5 M, 0.75 M, 1 M, and 1.5 M NaCl in buffer B (four fractions each). The triplex DNA-binding activity was recovered from the 1 M NaCl fractions. Starting from 2.5×10^{10} cells (≈100 g) of HeLa cells, the yields at each step were as follows: S-100, 333.9 mg (100%); S Sepharose, 137.1 mg; ammonium sulfate precipitation, 41.6 mg; hydroxylapatite, 7.76 mg; and triplex DNA Sepharose, 0.2 mg (0.06%).

Preparation of Triplex DNA Sepharose Column. One gram of CNBr-activated Sepharose 4B resin was coupled with 600 μg of poly(dA)·poly(dT) (≈300 nucleotides long) according to the supplier's manual. After blocking the reaction with 1 M ethanolamine hydrochloride (pH 8.0), the resin was washed with TMN buffer (10 mM Tris·HCl, pH 8.0/10 mM $MgCl_2$/50 mM NaCl) and suspended in 4 ml of the same buffer. The resin was then incubated at 37°C for 1 hr in the presence of 300 μg of poly(dT); this was followed by extensive washing with TMN buffer at room temperature and storage at 4°C. Formation of triplex DNA on the resin was monitored by the incorporation of radioactive $(dT)_{34}$ as a third strand. In a separate experiment single-strand oligonucleotides with a random sequence showed no incorporation. We estimate the total amount of DNA on the resin to be not more than 200 μg/ml of resin.

Gel Shift Assay. ^{32}P-labeled triplex or duplex DNA (1 nM, 20 μl) was incubated at 25°C for 30 min with proteins in 10 mM Tris·HCl, pH 8.0/10 mM $MgCl_2$/50 mM NaCl/1 mM dithiothreitol/0.05% (wt/vol) Nonidet P-40 in the presence or absence of a competitor DNA. The sample was mixed with 5% (wt/vol) glycerol/0.1% (wt/vol) bromophenol blue and then electrophoresed in a 4% polyacrylamide gel (mono:bis = 19:1) in HTM buffer (10 mM Hepes/10 mM Tris·HCl/5 mM $MgCl_2$, pH 8.0) at 10 V/cm for 2.5 hr at 4°C. Duplex or triplex DNA substrate was made by mixing 1 nM ^{32}P-labeled $(dT)_{34}$ and 100 nM $(dA)_{34}$ or 1 nM ^{32}P-labeled $(dA)_{34}$ and 100 nM $(dT)_{34}$, respectively; this was followed by incubation at 37°C for 30 min in HTM buffer. $(dT)_{34}$ or $(dA)_{34}$ was labeled with T4 polynucleotide kinase to a specific activity of 1.8×10^8 cpm/μg.

Purification of Triplex DNA. The triplex DNA substrate [1 nM ^{32}P-labeled $(dA)_{34}$ plus 100 nM $(dT)_{34}$] in HTM buffer was electrophoresed at 13 V/cm for 2.5 hr at 4°C in a 10% polyacrylamide gel in TBM buffer (89 mM Tris borate, pH 8.3/5 mM $MgCl_2$) and recovered from the gel by electroelution using Elutrap (Schleicher & Schuell) overnight at 100 V in TBM buffer; this was followed by dialysis in HTM buffer for 2 hr at 4°C.

RESULTS

Triplex DNA Formation. An oligo(dT)·oligo(dA)·oligo(dT) triplex (TAT triplex) can be formed by annealing an oligonucleotide $(dT)_{34}$ to a duplex plasmid DNA (pAT1) into which a $(dA)_{34}$·$(dT)_{34}$ duplex has been cloned. The triplex can be formed with either linear or covalently closed circular duplex DNA (Fig. 1*B*, lane 2 for supercoiled DNA and lane 6 for linear DNA) but cannot be formed if the third-strand oligonucleotide is the purine strand $(dA)_{34}$ (lanes 1 and 5). This triplex formation is dependent upon the presence of the AT duplex target sequence (lanes 3, 4, and 7–10).

To purify a protein that binds to this triplex we devised a gel shift assay based on the binding of a protein(s) to an oligonucleotide triplex. We established conditions for the formation of such an intermolecular oligonucleotide triplex (Fig. 2). Either labeled $(dA)_{34}$ or labeled $(dT)_{34}$ was titrated without (lane 1) or with (lanes 2–6) 1–10^4 molar ratios of the complementary oligonucleotides (Fig. 2 *A* or *B*, respectively). As observed with plasmid pAT1 (Fig. 1*B*), there was no detectable formation of triplex DNA with an excess of $(dA)_{34}$ (Fig. 2*B*, lanes 3–6), whereas triplex DNA did form with an excess of $(dT)_{34}$ (Fig. 2*A*, lanes 3–6). This triplex DNA formation is magnesium ion dependent (Fig. 2*C*). We estimate that the dissociation of the triplex DNA $(dT)_{34}$·$(dA)_{34}$·$(dT)_{34}$ to the duplex and single strand occurs at ≈20 nM.

Identification of a Triplex DNA-Binding Activity. Fig. 3 shows a gel shift assay using a triplex DNA substrate (see Fig. 2*A*, lane 4) and a duplex DNA substrate (see Fig. 2*B*, lane 4). These two substrates were used as a pair to search for a triplex DNA-specific binding protein present in nuclear extracts of HeLa cells. In Fig. 3*A*, nuclear protein fractions eluted from a hydroxylapatite column were assayed with the triplex (T, lanes 1–6) and the duplex (D, lanes 7–12) substrates. In addition to several bands representing nucleoprotein species derived from proteins with an affinity for both substrates, one nucleoprotein species (arrowhead in lane 5) was observed that appeared to be the product of a specific interaction between one protein(s) and only the triplex substrate. This triplex-binding protein(s) was purified further by

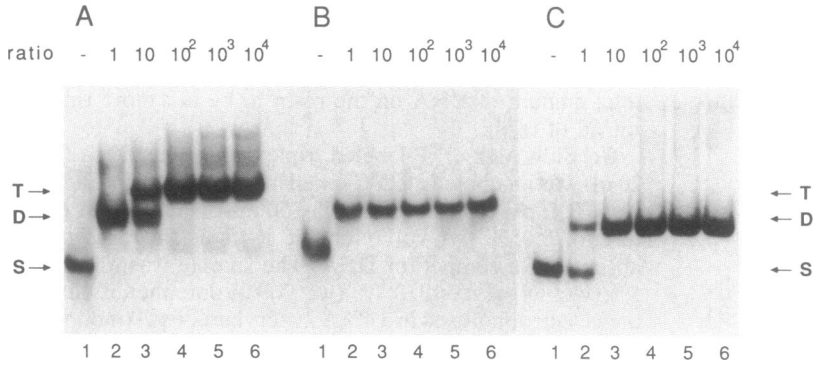

FIG. 2. Formation of an oligonucleotide triplex DNA, $(dT)_{34}\cdot(dA)_{34}\cdot(dT)_{34}$. One nanomolar ^{32}P-labeled $(dA)_{34}$ (*A* and *C*) or $(dT)_{34}$ (*B*) was incubated (37°C, 30 min) without (lane 1) or with (lanes 2–6) various amounts of the unlabeled complementary oligonucleotides in the presence of 5 mM $MgCl_2$ (*A* and *B*) or 5 mM EDTA (*C*). After the incubation, samples were electrophoresed on a 10% polyacrylamide gel in TBM buffer (*A* and *B*) or TBE buffer (*C*) at 4°C and autoradiographed. Positions of triplex (T), duplex (D), and single-strand (S) DNA are shown (Figs. 2–6).

triplex DNA Sepharose affinity chromatography (Fig. 3*B*). A protein (or proteins) with an affinity for the triplex substrate was eluted with 1 M NaCl (lane 5). This fraction showed no detectable signal with the duplex substrate (lane 11). We used this affinity-purified fraction for the characterization of the binding protein.

First, we confirmed that the nucleoprotein complex contains an intact triplex DNA (Fig. 4). The nucleoprotein species generated in the gel shift assay (lane 1) was excised and the bound DNA recovered by electroelution was electrophoresed in TBM buffer. The samples without and with 0.1% SDS (lanes 2 and 3) contain an intact triplex DNA. This result not only suggests that the nucleoprotein complex dissociates during electroelution but, more importantly, also indicates that the protein did indeed form a complex with triplex DNA. From this experiment we conclude that the nucleoprotein species generated from triplex DNA consists exclusively of triplex DNA and proteins and does not contain duplex or free single strands.

Next, we reevaluated the binding specificity using a gel-purified triplex DNA (radiolabeled dA strand) for the binding assay and compared the affinity for this triplex to that for pure duplex DNA and single-strand oligonucleotide $(dA)_{34}$ (Fig. 5). Both of these triplex and duplex substrates in Fig. 5 are distinct from the T and D substrates used above, in that those substrates were triplex with an excess of $(dT)_{34}$ (T) and duplex with an excess of $(dA)_{34}$ (D). In the presence of the purified triplex DNA and increasing quantities of the triplex DNA-binding protein, the electrophoretic mobility of the nucleoprotein species observed for triplex DNA and the protein remained unchanged (Fig. 5, lanes 2–4). Conversely, a nucleoprotein species with an identical electrophoretic mobility appeared with the duplex substrate when excess $(dA)_{34}$ was omitted (lanes 6–8), although the intensity of the band is 2- to 5-fold lower (average of several experiments) than that with the triplex DNA substrate. The protein fraction did not give rise to a major nucleoprotein species when the single-strand oligonucleotide $(dA)_{34}$ was used as the substrate (lanes 10–12). This last result indicates that the protein has little or no affinity for single-strand $(dA)_{34}$. Identical results were obtained with the T and D substrates when the radiolabel was on the dT strand and with single-strand $(dT)_{34}$ (data not shown). The low affinity of the protein for single-strand oligo(dT) was also observed in a competition experiment (see below). The low affinity of the protein for either single strand indicates that if there is a single-strand moiety on the triplex DNA it is not responsible for the binding observed.

The difference in the affinities of the protein for the triplex and duplex DNA is best demonstrated by examining the elution profiles of the protein from triplex and duplex DNA

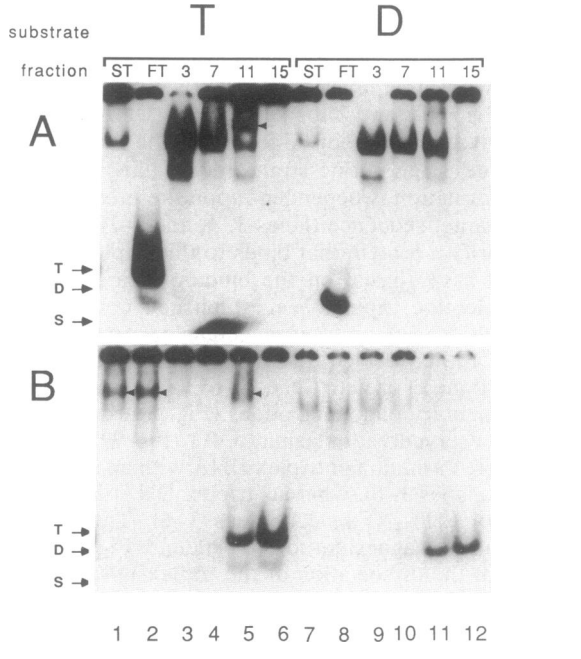

FIG. 3. Gel shift assay using an oligonucleotide triplex DNA. Protein fractions were incubated with triplex (T) substrate [1 nM ^{32}P-labeled $(dA)_{34}$, 100 nM $(dT)_{34}$, lanes 1–6] or duplex (D) substrate [1 nM ^{32}P-labeled $(dT)_{34}$, 100 nM $(dA)_{34}$, lanes 7–12] at 25°C for 30 min. Then the samples were electrophoresed on a 4% polyacrylamide gel in HTM buffer at 4°C for 2.5 hr and autoradiographed. (*A*) Hydroxylapatite column. ST, starting sample; FT, flow-through sample. Concentrations of the phosphate buffer are 200 mM (fraction 3), 300 mM (fraction 7), 400 mM (fraction 11), and 500 mM (fraction 15). (*B*) Triplex DNA Sepharose column. NaCl concentrations are 500 mM (fraction 3), 750 mM (fraction 7), 1 M (fraction 11), and 1.5 M (fraction 15). The arrowheads show the triplex DNA-specific bands.

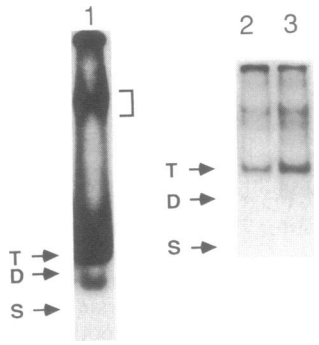

FIG. 4. The protein–triplex DNA complex contains an intact triplex DNA. Affinity-purified triplex DNA-binding protein fraction was used for the binding assay. After the gel shift assay (lane 1), the nucleoprotein species was excised (area in bracket) and the recovered material was electrophoresed (lane 2) or electrophoresed after mixing with 0.1% SDS (lane 3) on a 10% polyacrylamide gel in TBM buffer. The material migrating faster than triplex in lane 1 is not duplex.

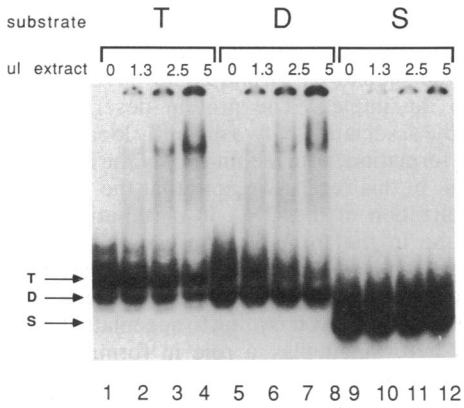

FIG. 5. Affinity of the triplex DNA-binding protein for a triplex, duplex, or single-strand DNA. Affinity-purified triplex DNA-binding protein was incubated with 1 nM gel-purified triplex (T), duplex (D), or single-strand (S) DNA with a ^{32}P-labeled dA strand in the presence of 100 nM control oligonucleotides (34-mer). After the incubation (25°C, 30 min), samples were electrophoresed on a 4% polyacrylamide gel in HTM buffer.

affinity columns (Fig. 6). Fractions from a TAT triplex DNA Sepharose column (Fig. 6*A*) and an AT duplex DNA Sepharose column (Fig. 6*B*) were assayed with the T substrate described above. As seen in Fig. 6, the protein was eluted with 1 M NaCl (fractions 9–12) in the triplex DNA column, whereas the elution from the duplex column was achieved at a lower salt concentration (750 mM NaCl; fractions 5–8). We cannot rule out an even more favorable situation—that is, that two different proteins are eluting from the two columns and that neither has an appreciable affinity for the other substrate.

Affinity of Triplex DNA-Binding Protein. Competition experiments to ascertain the affinity of the triplex DNA-binding protein for the TAT triplex were performed using several competitor DNAs (data not shown). First, using gel-purified triplex we observed that the binding of the protein was not affected even in the presence of a 500-fold molar excess of $(dT)_{34}$; this result suggests that the protein has a very low affinity for single-strand oligo(dT). Next, triplex DNA-binding protein was added to 1 nM ^{32}P-labeled $(dA)_{34}$ and 100

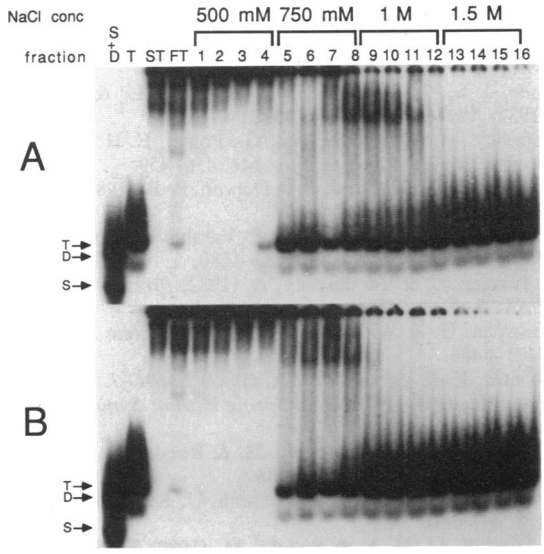

FIG. 6. Affinity of the triplex DNA-binding protein for triplex or duplex DNA affinity column. Hydroxylapatite column fractions were applied to a triplex (*A*) or a duplex (*B*) DNA Sepharose column (bed vol, 1 ml; fraction vol, 0.5 ml). Fractions were assayed with T substrate. ST, starting sample; FT, flow-through sample.

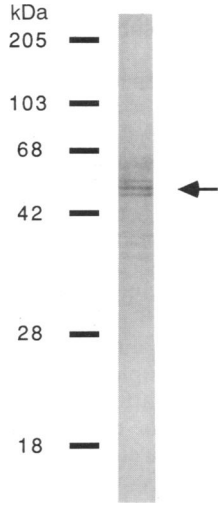

FIG. 7. Identification of the triplex DNA-binding protein. Twenty microliters of the affinity-purified protein fraction was electrophoresed on an SDS/polyacrylamide gel (12% acrylamide). Major protein species at 55 kDa are indicated (arrow).

nM $(dT)_{34}$ in the presence of several other competitor DNAs. Many DNAs [*E. coli* DNA, human DNA, poly(dA-dC)·poly(dT-dG), and poly(dG)·poly(dC)] competed very weakly (that is, not at all in the presence of a 50-fold molar excess of the competitor DNAs and somewhat in the presence of a 500-fold molar excess) or not at all even in the presence of a 500-fold molar excess [poly(dG-dC)·poly(dC-dG) and poly(dA-dT)·poly(dT-dA)]. Two DNAs, however, poly(dA)·poly(dT) and poly(dA-dG)·poly(dT-dC), were better competitors. Both of these DNAs competed well at a 50-fold molar excess and abolished all binding at a 500-fold molar excess. The affinity for the AT duplex is also seen in Fig. 5. However, poly(dA-dT)·poly(dA-dT), an isomer of poly(dA)·poly(dT), did not compete for the binding of this protein to the triplex.

Molecular Characterization of the Purified Triplex DNA-Binding Protein. The SDS/polyacrylamide gel electrophoretic pattern of the affinity-purified triplex DNA-binding protein (Fig. 7) shows three species of proteins with apparent molecular mass of ≈55 kDa. We confirmed the triplex DNA binding activity of these protein species by two independent methods (data not shown). (*i*) UV cross-linking of the protein with radioactive triplex DNA followed by pancreatic DNase treatment and SDS/polyacrylamide gel electrophoresis revealed two of these 55-kDa bands to be the major species that were bound to the triplex DNA. (*ii*) Iodination with ^{125}I of the proteins recovered from the purified nucleoprotein complex again revealed two of the same 55-kDa proteins (data not shown).

DISCUSSION

Triplex DNA Binding. To our knowledge, a protein with an affinity for triplex DNA has not been reported previously. To demonstrate the specificity of the protein for $(dT)_{34}$·$(dA)_{34}$·$(dT)_{34}$ triplex (TAT triplex) and to obtain more information about the binding protein, we did the following experiments. We first observed that the binding occurs between the protein and triplex DNA in the presence of a 98-fold excess of $(dT)_{34}$ (Fig. 4). Second, the same nucleoprotein species was observed when a purified triplex DNA substrate was used, and the affinity for this purified triplex was shown to be higher than that for the duplex DNA or the constituent single-strand oligonucleotides (Fig. 5). Third, the elution profiles from a triplex and a duplex DNA column also confirmed the higher affinity of the protein for the TAT triplex over the AT duplex (Fig. 6). The low affinity of the protein for single-strand $(dA)_{34}$ and $(dT)_{34}$ precludes that the binding to the triplex DNA is mediated by putative single-strand tails on the

triplex. Finally, the competition experiments confirmed the specificity of the protein.

Although we found no discrepancy between the results with and without excess oligo(dT) for the triplex DNA binding, the duplex substrates with (Fig. 3 *A* and *B*, lane 11) and without excess oligo(dA) (Fig. 5, lanes 6–8) showed a marked difference: a retarded band appeared with the duplex at the identical position as with the triplex substrate but only in the absence of excess oligo(dA). Perhaps the protein facilitates the disproportionation of the duplex into triplex and single strand and the protein binds to the resulting triplex. In the presence of an excess oligo(dA) such a disproportionation would be highly unfavorable. Formally, another possibility is that the binding of the protein to duplex is competitively inhibited by excess oligo(dA). This latter possibility is unlikely in view of the low affinity of the protein for single-strand oligo(dA) (Fig. 5, lanes 9–12). In a similar vein, we cannot rule out that the molecular species in the poly(dA)·poly(dT) competition experiment described above was not the duplex but some triplex formed in the course of the experiment. In summary, it is quite possible that, as the experiments in Fig. 3 would seem to indicate, the protein does not have an appreciable affinity for the AT duplex.

Protein Identification. Some known proteins, such as histones H2A (23) and H1 (26), datin (23), and topoisomerase II (16) also bind to (A+T)-rich sequences. The molecular mass of the protein we have purified is different from that of these other proteins. In addition, although we found that histones also bind to the triplex DNA substrate, under our assay conditions, when histones are added to the TAT triplex they do not give rise to a distinct nucleoprotein species but form an aggregate at the top of the gel (data not shown). Furthermore, we could not detect topoisomerase II activity in our purified protein preparation (data not shown).

Protein–DNA Interaction. What is being recognized by this TAT triplex binding protein? Interestingly, many of the AT duplex binding proteins bind to DNA in the minor groove (27, 28), whereas most of sequence-specific DNA-binding proteins such as transcription factors bind to DNA in the major groove. Since the major groove is occupied by the third-strand dT it is unlikely that the protein is forming hydrogen bonds with the hydrogen bond donors or acceptors on the duplex bases in the major groove. Since the protein also exhibits a somewhat lower affinity for the corresponding duplex the protein could interact with any number of other structural features that are qualitatively similar in the TAT triplex and the AT duplex. A somewhat trivial quantitative difference between the duplex and triplex is their difference in charge density and this might contribute to the difference in affinities observed. On the other hand, electrostatic interactions between the protein and the DNA phosphates cannot account for a major component of the binding energy because the protein has a low affinity for most duplex and single-strand DNAs. A more interesting possibility is that the protein forms hydrogen bonds to the duplex bases in the minor groove and that these interactions are quantitatively enhanced in the triplex—that is, the third strand is not directly involved in the binding yet in the process of annealing to the duplex the third strand indirectly enhances the binding of the protein by altering the minor groove (2).

Insight into the nature of the recognition by this protein awaits elucidation of the specific contacts made by the protein with the triplex and duplex DNAs.

Biological Significance of the Triplex DNA-Binding Protein. The 2- to 5-fold higher affinity of the protein for the triplex over the duplex DNA (Fig. 5) prompted us to investigate whether the protein might promote triplex formation from a starting duplex and single-strand DNA. Preliminary data, using the most purified fraction, showed that the protein enhanced the triplex formation (data not shown); this could be the function of the protein *in vivo*. We note that, as AT duplex DNA is preferentially excluded from nucleosomes (29, 30), the second dT strand could have relatively facile access to the duplex. The protein described here might facilitate the association of two distantly located AT duplexes by triplex formation, with looping out of the intervening DNA sequences. In this regard we note that the subnuclear compartmentalization of the DNA in chromatin has been proposed to be in the form of loops containing from a few kilobase pairs to >100 kilobase pairs and anchored to the nuclear scaffold by highly (A+T)-rich regions called scaffold-associated regions. It is tempting to speculate that the protein described here might play a role in forming these looped chromatin domains.

In order to elucidate the biological meaning of triplex DNA-binding proteins, it might be useful to study the interaction between triplex DNA and other proteins with a known biological function.

We thank Carol Camerini-Otero, Lance Ferrin, Moon Kim, Peggy Hsieh, and other members of the Genetics and Biochemistry Branch for comments on the manuscript and helpful discussions; George Poy for oligonucleotide syntheses; and Linda Robinson for her assistance.

1. Felsenfeld, G., Davies, D. R. & Rich, A. (1957) *J. Am. Chem. Soc.* **79**, 2023–2024.
2. Arnott, S. & Selsing, E. (1974) *J. Mol. Biol.* **88**, 509–521.
3. Pilch, D. S., Levenson, C. & Shafer, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1942–1946.
4. Fox, K. R. (1990) *Nucleic Acids Res.* **18**, 5387–5391.
5. Morgan, A. R. & Wells, R. D. (1966) *J. Mol. Biol.* **37**, 63–80.
6. Moser, H. E. & Dervan, P. B. (1987) *Science* **238**, 645–650.
7. Mirkin, S. M., Lyamichev, V. I., Drushiyak, K. N., Dobrynin, V. N., Filippov, S. A. & Frank-Kamenetskii, M. D. (1987) *Nature (London)* **330**, 495–497.
8. Kohwi, Y. & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.
9. Hsieh, P., Camerini-Otero, C. S. & Camerini-Otero, R. D. (1990) *Genes Dev.* **4**, 1951–1963.
10. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988) *FASEB J.* **2**, 2939–2949.
11. Hentchel, C. C. (1982) *Nature (London)* **295**, 714–716.
12. Nickol, J. M. & Felsenfeld, G. (1982) *Cell* **35**, 467–477.
13. Behe, M. J. (1987) *Biochemistry* **26**, 7870–7875.
14. Beasty, A. M. & Behe, M. J. (1988) *Nucleic Acids Res.* **16**, 1517–1528.
15. Gasser, S. M. & Laemmli, U. K. (1986) *Cell* **46**, 521–530.
16. Adachi, Y., Kas, E. & Laemmli, U. K. (1989) *EMBO J.* **8**, 3997–4006.
17. Lee, J. S., Latimer, L. J. P., Haug, B. L. & Braun, R. P. (1987) *Nucleic Acids Res.* **15**, 1047–1061.
18. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. (1988) *Science* **241**, 456–459.
19. Maher, L. J., III, Wold, B. & Dervan, P. B. (1989) *Science* **245**, 725–730.
20. Solomon, M. J., Strauss, F. & Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1267–1280.
21. Levinger, L. & Varshavsky, A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7152–7156.
22. Garreau, H. & Williams, J. G. (1983) *Nucleic Acids Res.* **11**, 8473–8484.
23. Winter, E. & Varshavsky, A. (1989) *EMBO J.* **8**, 1867–1877.
24. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
25. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1488.
26. Kas, E., Izaurralde, E. & Laemmli, U. K. (1989) *J. Mol. Biol.* **210**, 587–599.
27. Churchill, M. E. A. & Suzuki, M. (1989) *EMBO J.* **8**, 4189–4195.
28. Reeves, R. & Nissen, M. S. (1990) *J. Biol. Chem.* **265**, 8573–8582.
29. Simpson, R. T. & Kunzler, P. (1979) *Nucleic Acids Res.* **6**, 1384–1415.
30. Rhodes, D. (1979) *Nucleic Acids Res.* **6**, 1805–1816.