Protein A1 (M, ~32,000), a major glycine-rich protein of heterogeneous nuclear ribonucleoproteins (hnRNP), was purified to near homogeneity under nondenaturing conditions from HeLa cells. Limited proteolysis of the native protein yields a trypsin-resistant N-terminal nucleic acid-binding domain about 195 amino acids long which has a primary structure nearly identical to that of the 195-amino acid-long single-stranded DNA (ssDNA)-binding protein UP1 (M, 22,162) from calf thymus (Williams, K. R., Stone, K. L., LoPresti, M. B., Merrill, B. M., and Planck, S. R. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 5666–5670). Of the 61 glycine residues of A1 are present in the trypsin-sensitive C-terminal domain of the protein which contains no sequences homologous to UP1. Protein A2, another major glycine-rich core hnRNP protein from HeLa, has a domain structure analogous to A1 and appears to be related to ssDNA-binding proteins UP1-B from calf liver and HDP-1 from mouse myeloma in a way similar to the A1/UP1 relationship. In contrast to ssDNA-binding proteins, A1 binds preferentially to RNA over ssDNA and exhibits no helix-stabilizing activity.

In eukaryotic cells, the newly synthesized hnRNA associates with a set of proteins and forms ribonucleoprotein particles (hnRNP). These particles can be seen in electron micrographs of transcriptionally active chromatin as fibrils containing 20–30 nm beads spaced along their exposes, generally believed to be part of the hnRNA-processing apparatus (for review, see Refs. 1 and 2). The 20–30 nm substructures, often called monocaplets, can be recovered from purified nuclei as a fairly homogenous peak sedimenting at about 30–40 S in the sucrose gradient. The monocaplets which arise during isolation as a result of endocytotic cleavage of the larger hnRNP complexes contain, in addition to pre-mRNA sequences, a set of about nine prominent polypeptides called core proteins and a number of minor protein components (4). In HeLa cells, the major core proteins are termed A1, A2, B1 (two or three subspecies), B2, C1, C2, and C3, in order of increasing molecular weights (4). Most of the core proteins are modified post-translationally, primarily by phosphorylation. The number of core proteins in a variety of organisms is similar, and they all have molecular weights between 32,000 and 42,000. Most are basic proteins that have similar amino acid compositions characterized by a high content of glycine (about 30%), a low content of cysteine, the presence of the unusual modified amino acid dimethylarginine, and a blocked N terminus. Immunological investigations, including the use of monoclonal antibodies, demonstrate that core proteins comprise a family of related and evolutionarily conserved polypeptides (5, 6). Furthermore, a cDNA clone of a major hnRNP protein from brine shrimp Artemia salina cross-hybrids with DNA sequences of plant, avian, and mammalian origin (7). The core proteins appear to be devoid of any enzymatic activity and are believed to function as structural components that condense and organize hnRNA into the characteristic beads-on-a-string hnRNP particles. Nucleoprotein complexes that resemble native hnRNP can be reconstituted from exogenous RNA and these proteins (8–11). It has been shown by optical methods that the RNA in hnRNP, both native and reconstituted, is substantially devoid of any residual secondary structure, indicating that at least some of the core proteins may have RNA helix-stabilizing properties (12–14). A major hnRNP protein that was purified to homogeneity under nondenaturing conditions from A. salina was shown to be an RNA helix-stabilizing protein (15, 16).

It was recently demonstrated that the ssDNA-binding protein (also termed a DNA helix-stabilizing protein) from calf thymus, UP1 (17), shares some antigenic determinants in common with core hnRNP proteins from calf thymus and HeLa cells (18). Proteins analogous to UP1 were also isolated from mouse myeloma (19). In calf thymus, mouse myeloma, and probably in HeLa cells (18), there appears to be a group of proteins related structurally and antigenically to UP1 (UP1-type proteins) with molecular weights ranging from 22,000 to 28,000, and immunological investigations suggest that these proteins may arise by the action of an endogenous protease on the larger and more abundant core hnRNP proteins (20), e.g. the mouse myeloma UP1-type protein obtained by standard purification has a M, of about 24,000, whereas an immunoblot of the crude homogenate detects a species of M, ~36,000 (21). Furthermore, a full-length cDNA clone of a UP1-type protein from rat was found to contain an open reading frame coding for a protein of 320 rather than the expected ~200 amino acids (22). The clone was isolated by screening a cDNA library constructed with total poly(A)* RNA of newborn rat brain with oligonucleotide probes based on peptide sequences of calf thymus UP1. The complete
sequence of the 195 amino acids of calf thymus UP1 (M,
22,162) has been determined (23) and was found to be iden-
tical to the predicted sequence of the 195-amino acid-long N-
terminal fragment of the rat clone (22). Partial amino acid
sequencing of an analogous protein from mouse myeloma also
shows a high degree of sequence homology with calf thymus
UP1 (23). The UP1-type proteins bind tightly to single-
stranded but not double-stranded DNA and stimulate the
activity of the homologous DNA polymerase α (24, 25). They
also have a substantial affinity for single-stranded RNA.
These properties of a UP1-type mammalian proteins are
similar to those of the well-characterized prokaryotic ssDNA-
binding proteins, the phage T4 gene 32 protein, and the
Escherichia coli SSB DNA protein that play essential roles in
replication, recombination, and repair. However, a fundamen-
tal property of prokaryotic ssDNA-binding proteins, and one
thought to be critical for their function, viz. cooperative
binding to single-stranded DNA, is not shared by UP1-type
proteins.

In this report, we describe the purification of core proteins
A1, B2, C1, and A2 from HeLa cells, the first three under
nondenaturing conditions. We analyze the domain structure
of A1 and A2 and their relationship to some ssDNA-binding
proteins.

EXPERIMENTAL PROCEDURES AND RESULTS

Primary Structures of the N-terminal Fragment of HeLa
Core Protein A1 and the ssDNA-binding Protein UP1 from
Calf Thymus Are Nearly Identical—Limited proteolysis of a
native protein often generates large polypeptides that may
give some insights into the domain structure of the protein.
As seen from Fig. 4, partial digestion of A1 (M, ~32,000) with
trypsin at 0 °C yields, within a few minutes, a fragment of M,
~22,000 that remains completely resistant to further prote-
olysis for at least 2 h, i.e. for about 1 h after the disappearance
of the A1 substrate. In contrast, the remaining part of the A1
molecule, about 10,000–11,000 daltons, is sensitive to trypsin
and seems to be completely degraded to small peptides that
run off the gel. The M, 22,000 tryptic fragment of A1 appears
to have a blocked N-terminal amino acid since no discernible
sequence was obtained when the material, electroeluted from
SDS-polyacrylamide gels, was subjected to gas-phase sequenc-
ing. Protein A1 and other major core proteins apparently have
blocked N termini, thus locating the M, 22,000 tryptic frag-
ment at the N-terminal end of A1. The trypsin-derived frag-
ment of A1 migrates on the SDS-polyacrylamide gel at the
same rate as authentic calf thymus UP1 (Fig. 4, lane 2) and
has a similar amino acid composition (data not shown). Since
calf thymus UP1 also has a blocked N-terminal amino acid,
these results prompted us to further examine and compare
these two proteins. The complete primary structure of UP1
has been recently determined (23); it contains 195 amino
acids (M, 22,162), has 2 cysteines, and contains a single
N2,N3-dimethylarginine residue near its C terminus. A com-
parison of HPLC maps of peptides from exhaustive tryptic
digests of A1 and UP1 (Fig. 5) demonstrates that virtually
every UP1 tryptic peptide makes its appearance at the same
position in the A1 map. The A1 pattern shows some 10–11
additional peptides absent from the UP1 map (indicated by
arrows in Fig. 5) which are presumably derived from the
tryptsin-sensitive C-terminal fragments of the protein. There
are about 43 peptides in the HPLC map of A1, in fairly good
agreement with the number of lysine and arginine residues in
the protein (Table I). The very high degree of homology
between UP1 and the N-terminal domain of A1 indicated by
the similarities of their HPLC maps has been confirmed by
sequencing of six tryptic peptides of A1, four of which match
the sequence of UP1 (Figs. 6 and 7). Of a total of 46 amino
acids sequenced in the latter four peptides, 45 correspond
exactly to sequences found in UP1. The one substitution
involves the interchange of lysine in A1 for arginine in UP1
at position 30 (Fig. 6). The four matching A1 peptides cover
UP1 sequences near the N terminus (positions 14–30), in the
middle of the protein (positions 92–104), and near the C
terminus (positions 146–153 and 183–194) of the protein.
This, together with the HPLC data of Fig. 5, provides con-
vincing evidence that the N-terminal M, 22,000 fragment of
HeLa A1 has a primary structure nearly identical to that of
calf thymus UP1.

UP1 is not a glycine-rich protein; it contains 7.7 mol %
glycine, which is nearly identical to the 7.6 mol % frequency
of glycine in the 2054 eukaryotic protein sequences in the
PIR Protein Sequence Database. In contrast to UP1, A1

---

2 Portions of this paper (including "Experimental Procedures," part
of "Results," and Figs. 1–3) are presented in miniprint at the end of
this paper. Miniprint is easily read with the aid of a standard
magnifying glass. Full size photocopies are available from the Journal
of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814.
Request Document No. 86M-360, cite the authors, and include a
check or money order for $3.20 per set of photocopies. Full size
photocopies are also included in the microfilm edition of the Journal
that is available from Waverly Press.
peptides that come from the Cterminus of the protein and procedures.

calf liver UP1-B (0.7 nmol) proteins liver UP1-B profiles. The large absorbance peak at about 23 min is due to Tris-HCl.

cetic acid-precipitated proteins were digested in 2

cine-rich (Table I). In view of the near identity of UP1 and the calf thymus UP1 (0.9 nmol) proteins from the respective calf thymus UP1 (0.9 nmol) proteins (top) or from the HeLa A2 (1.0 nmol) and the calf liver UP1-B (0.7 nmol) proteins (bottom). The trichloroacetic acid-precipitated proteins were digested in 2 M, Ty r 13.2 4 Met 4.7 4 Ala 13.6 10 Gl~Y 60.5 15 Glx 28.4 25 Thr 10.7 12 Asx 33.1 17 Arg 22.3 15

TABLE I

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No. of residues</th>
<th>UP1</th>
<th>A1</th>
<th>A2</th>
<th>UP1-B</th>
<th>A1/UP1</th>
<th>A2/UP1-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asx</td>
<td>33.1</td>
<td>17</td>
<td>16</td>
<td>33.7</td>
<td>18.1</td>
<td>15.6</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>10.7</td>
<td>12</td>
<td>0</td>
<td>9.8</td>
<td>10.9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>26.0</td>
<td>16</td>
<td>10</td>
<td>21.7</td>
<td>14.6</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>Gix</td>
<td>28.4</td>
<td>25</td>
<td>3</td>
<td>31.2</td>
<td>28.3</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>11.7</td>
<td>6</td>
<td>6</td>
<td>14.6</td>
<td>5.5</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>60.5</td>
<td>15</td>
<td>45</td>
<td>70.0</td>
<td>21.7</td>
<td>48.3</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>13.6</td>
<td>10</td>
<td>4</td>
<td>13.6</td>
<td>11.5</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>15.4</td>
<td>17</td>
<td>0</td>
<td>13.9</td>
<td>14.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>4.7</td>
<td>4</td>
<td>1</td>
<td>5.2</td>
<td>3.4</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>7.7</td>
<td>8</td>
<td>0</td>
<td>8.4</td>
<td>7.8</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>10.0</td>
<td>8</td>
<td>2</td>
<td>11.2</td>
<td>8.6</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>13.2</td>
<td>4</td>
<td>9</td>
<td>18.0</td>
<td>4.4</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>17.5</td>
<td>10</td>
<td>8</td>
<td>16.6</td>
<td>11.6</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>6.7</td>
<td>8</td>
<td>0</td>
<td>7.1</td>
<td>7.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>18.4</td>
<td>16</td>
<td>2</td>
<td>16.3</td>
<td>14.5</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>22.3</td>
<td>17</td>
<td>7</td>
<td>26.4</td>
<td>16.7</td>
<td>9.7</td>
<td></td>
</tr>
<tr>
<td>Me-Arg*</td>
<td>3.1</td>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M,</td>
<td>32,000*</td>
<td>22,162</td>
<td>34,000</td>
<td>22,000*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Me-Arg, dimethylarginine.

Table I shows the amino acid composition of HeLa proteins A1 and A2 and calf proteins UP1 (thymus) and UP1-B (liver) and the number of residues in each peptide. The table is used to determine the number of glycine residues in each peptide.

The HPLC map of A1 that are absent from the respective calf thymus UP1 and calf liver UP1-B profiles. The large absorbance peak at about 23 min is due to Tris-HCl.

contains 20.0 mol % glycine, which makes it unusually glycine-rich (Table I). In view of the near identity of UP1 and the N-terminal domain of A1, one would expect that those A1 peptides that come from the C terminus of the protein and do not match UP1 sequences should contain a high number of glycine residues. In fact, since the calculated mass of the C-terminal fragment of A1 is about 10,000–11,000 daltons

(see Fig. 4 and Table I), its glycine content should be nearly 40 mol % (Table I). This prediction appears to be borne out by the sequences of two additional A1 tryptic peptides that show no homology to UP1 (Fig. 7). Of a total of 33 amino acids so far sequenced in these peptides, 15 (45%) are glycine residues. Also note from Table I that the calculated number of lysine and arginine residues in the C-terminal fragment of A1 is in fairly good agreement with the number of peptides in the HPLC map of A1 that are absent from the HPLC map of UP1 (Fig. 5). Finally, the amino acid sequences of the two peptides in Fig. 7 have recently been shown to match exactly positions 197–215 and 301–315 of the full-length rat clone (Ref. 22; see text).
sequences of Fig. 7, the clone appears to correspond to a rat A1-type protein.

**Tryptic Digests of Core Proteins A2, B2, and C1**—Protein A2 (M, ~34,000), a major glycine-rich core protein which is present in 40 S monoparticles at a molar ratio of approximately 1:1 to protein A1, was obtained pure by CM-cellulose chromatography in 6.0 M urea or electroelution from SDS-polyacrylamide gels (Fig. 3 and text). The amino acid compositions of A1 and A2 are similar (Table I; cf. Ref. 3). The HPLC map of peptides from an exhaustive tryptic digest of A2 (Fig. 5) contains about 41 peptides, in good agreement with the number of lysine and arginine residues in A2 (Table I). An HPLC map of tryptic peptides from an exhaustive digest of protein UP1-B (M, ~22,000) is shown in Fig. 5. This ssDNA-binding protein was isolated along with UP1 from calf liver according to the procedure for the purification of UP1; it has an amino acid composition similar to UP1 (Table I). In addition, UP1 and UP1-B share a high degree of sequence homology. Of the 103 amino acids that have so far been sequenced in UP1-B, 84 have been shown to be identical to sequences in UP1. A comparison of the peptide maps of A2 and UP1-B shows that essentially each of the UP1-B trypsin peptides has its counterpart which elutes at the same position in the A2 map. In addition, the A2 map contains about 11 peptides that are absent from the UP1-B pattern (indicated by arrows in Fig. 5), in good agreement with the calculated number of lysine and arginine residues which can be assumed to be present in the 10,000-11,000-dalton fragment(s) of A2 remaining after subtracting the molecular mass of UP1-B from that of A2 (Table I). In view of the similarities of the peptide maps of the two proteins, these are presumably the glycine-rich peptides since the glycine content of UP1-B and A2 is 18 and 70 residues, respectively.

Three A2 trypsin peptides have been sequenced so far. One of these sequences proved to match exactly residues 113-121 in UP1, whereas a second matched UP1 residues 166-172 with the exception that the A2 sequence has an isoleucine corresponding to valine 169 in UP1. The third peptide has a sequence, Glu-Glu-Ser-Gly-X-Pro-Gly-Ala-Val-Thr-Val, which is highly homologous to residues 92-103 in UP1 (Fig. 6) and is identical to the corresponding region in an analogous ssDNA-binding protein, HDP-1, isolated from mouse myeloma cells (23). Taken together with a previous study that indicates that the UP1 and UP1-B proteins may share as much as 82% sequence homology, these data indicate that the A1 and A2 hnRNP proteins are encoded by two different genes that nonetheless share a high degree of sequence homology. The UP1 and UP1-B ssDNA-binding proteins could result from limited proteolysis of the A1 and A2 hnRNP proteins, respectively (see "Discussion"). The HDP-1 ssDNA-binding protein that was previously reported in mouse myeloma cells (19) appears to result from cleavage of an A2 hnRNP protein. Thus, previous sequence differences between the calf thymus UP1 and mouse myeloma HDP-1 proteins result from the fact that these proteins are related to A1 and A2, respectively, and not, as was originally proposed (23), from interspecies sequence differences in UP1.

Proteins B2 and C1 were purified under nondenaturing conditions (see Miniprint) and subjected to partial trypsin digestion. As seen from Fig. 8, regardless of whether the proteins were digested at 22 or 37 °C, the results are analogous to those obtained with A1 (Fig. 4) in that each protein loses uniformly about 12,000-13,000 daltons and a large trypsin-resistant fragment remains intact. The remaining part of the C1 molecule appears to be completely degraded to small peptides that run off the gel as is the case with A1. Protein B2 (M, ~38,000) seems to have a single trypsin-sensitive bond under our conditions of digestion (enzyme-to-substrate ratio of 1:30, w/w) since it yields two fragments of M, ~26,000 and ~13,000.

**Binding of A1 to Polynucleotides**—We have examined some of the nucleic acid binding properties of A1, and we find that they differ considerably from those of UP1. We did not find any perturbation of CD spectra of HeLa hnRNA, coliphage MS2 RNA, poly[d(A-T)], poly[r(A+U)], and poly(rA) at 25-35 µM nucleotide upon the addition of A1 up to a molar ratio of one protein per 8-12 nucleotides. On the other hand, experiments employing the nitrocellulose filter binding assay (Table II) have shown that, at 10 µM nucleotide, A1 forms complexes with a variety of labeled natural and synthetic RNAs and single-stranded and double-stranded DNA. This suggests that, in contrast to UP1 (28), binding of A1 is not accompanied by conformational changes indicative of the unfolding of the secondary structure of the polynucleotide. Whereas the UV and CD spectra of the homopolymer pair (rA + rU) are not perturbed by A1, addition of the protein affects the rate and extent of duplex formation at low ionic strength. In a typical experiment shown in Fig. 9, the addition of A1 to poly(rU) in a buffer containing 20 mM NaCl at a ratio of one protein per 12 nucleotides prior to the addition of poly(rA) increases the t₁/₂ for duplex formation from 0.95 min in the absence of A1 to 4.1 min in the presence of A1. Essentially the same results are obtained regardless of

---

8 B. Merrill, K. Stone, S. Riva, and K. R. Williams, unpublished data.
The protein from the initial complex seems unlikely since binding assay was also used to measure the relative affinity of the poly[r(U+A)] duplex was 44%.

The binding of A1 to polynucleotides is sensitive to elevated NaCl. The nitrocellulose binding assay was as described under "Experimental Procedures" with protein A1 at 0.8 µM and polynucleotides at 10.0 µM (nucleotide). The extent of polynucleotide retention on nitrocellulose filters is reported in Table II.

![Fig. 9. Effect of protein A1 on the formation of the poly[r(U+A)] duplex. Poly[r(U)] (4.0 nmol) was mixed with poly[r(A)] (4.0 nmol) in the presence (curve 1) or absence (curve 2) of A1 (0.33 nmol) at 20°C in a buffer containing 10.0 mM Tris-HCl, pH 7.5, and 20 mM NaCl in a final volume of 0.25 ml. Changes in absorbance at 260 nm were recorded. The arrow indicates addition of NaCl from a 300 mM stock solution. The hyperchromicity of the random coil versus duplex was 44%.

Whether the protein is initially mixed with the polypurine or the polypyrimidine. The decrease in the rate of annealing suggests that some rearrangement of the initial A1-single-stranded polynucleotide complex is required to allow duplex formation and could be the rate-limiting step. Dissociation of the protein from the initial complex seems unlikely since under these conditions A1 binds to synthetic duplex RNA (Table II). Also note from Fig. 9 that the presence of A1 allows about 60% of the duplex to be formed at 20 mM NaCl. A somewhat higher NaCl concentration, which favors duplex formation, is required to complete the reaction.

The binding of A1 to polynucleotides is sensitive to elevated concentrations of NaCl (Table II). Whereas complexes of UP1 with single-stranded DNA are stable in buffers containing up to 0.4–0.5 M NaCl (17), dissociation of A1-poly nucleotide complexes starts at about 0.09–0.1 M NaCl (not shown) and is complete at 0.5 M NaCl (Table II). The nitrocellulose binding assay was also used to measure the relative affinity of A1 for various polynucleotides. Table III shows the results of competition experiments in which a 3-fold excess of unlabeled polynucleotide was mixed with HeLa [3H]hnRNA or with coliphage MS2 [3H]RNA (each at 10.0 µM nucleotide) prior to the addition of A1. The results indicate a preference for the binding of natural RNA since none of the synthetic polymeric nor DNA competed effectively for A1 with MS2 RNA or with hnRNA, whereas MS2 RNA reduced the binding to hnRNA by more than half.

**Table II**

<table>
<thead>
<tr>
<th>Polynucleotide retained on filter</th>
<th>NaCl concentration (M)</th>
<th>% input</th>
</tr>
</thead>
<tbody>
<tr>
<td>hnRNA</td>
<td>40 mM</td>
<td>100</td>
</tr>
<tr>
<td>Coliphage MS2 RNA</td>
<td>40 mM</td>
<td>100</td>
</tr>
<tr>
<td>Poly[r(U)]</td>
<td>40 mM</td>
<td>70</td>
</tr>
<tr>
<td>Poly[r(A)]</td>
<td>40 mM</td>
<td>69</td>
</tr>
<tr>
<td>Poly[r(A+U)]</td>
<td>40 mM</td>
<td>75</td>
</tr>
<tr>
<td>Poly[r(A+U)]</td>
<td>300 mM</td>
<td>76</td>
</tr>
<tr>
<td>Poly[r(A+U)]</td>
<td>500 mM</td>
<td>73</td>
</tr>
</tbody>
</table>

*ND, not determined; dsDNA, double-stranded DNA.

**Table III**

<table>
<thead>
<tr>
<th>Competing polynucleotide</th>
<th>[3H]hnRNA</th>
<th>MS2 [3H]RNA</th>
<th>% input</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>rU</td>
<td>72</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>rC</td>
<td>100</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>rA</td>
<td>95</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>rl</td>
<td>ND*</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>r(A+U) (1:1)</td>
<td>92</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>MS2 RNA</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ssDNA (calf thymus)</td>
<td>78</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>dsDNA (calf thymus)</td>
<td>90</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>d(A-T)</td>
<td>95</td>
<td>ND</td>
<td></td>
</tr>
</tbody>
</table>

*ND, not determined; dsDNA, double-stranded DNA.

**DISCUSSION**

The data presented in this paper show that the primary structure of the ssDNA-binding protein UP1 from calf thymus is nearly identical to that of the N-terminal domain of HeLa core hnRNP protein A1. The degree of homology, based on partial sequences of A1 shown in Fig. 6, is 98%, the only difference being the interchange of the chemically similar amino acids lysine and arginine. This latter difference may result from species-specific sequence differences between the HeLa cell A1 and the calf thymus UP1 protein. The sequence of the N-terminal 195 amino acids of the 320-amino acid-long presumptive UP1-type protein from rat was predicted from the sequence of the cDNA clone of Cobianchi et al. (22) illustrating that this is identical to that of calf thymus UP1. Since those HeLa A1 peptides from the C-terminal domain which have been sequenced (Fig. 7, a third of the domain) match exactly to the rat sequence that extends beyond the N-terminal domain, the clone most likely corresponds to a rat A1 protein. Although interspecies relatedness of A1 was inferred from chemical and immunological analysis of core proteins, such a high degree of homology was unexpected and is indicative of profound evolutionary constraints. Partial amino acid sequencing of a major hnRNP protein from A. salina also demonstrates considerable homology to UP1.4 Our analysis of protein A2 (Fig. 5 and sequences in text), taken together with sequence data on the ssDNA-binding protein UP1-B from calf liver and on the HDP-1 ssDNA-binding protein from mouse myeloma (23), indicates that these two proteins are related to A2 in a way analogous to the A1/UP1 relationship. These results suggest that at least some ssDNA-binding proteins, until recently considered a separate class, and some core hnRNP proteins could be products of the same genes. In A1 and UP1 is one example; and A2, UP1-B, and HDP-1 is another. In view of the extensive sequence homologies between UP1, UP1-B, and HDP-1, it is obvious that A1 and A2 are closely related proteins. In fact, as can be seen from Fig. 5, the two proteins share about 23 tryptic peptides of a total of about 43.

Whereas A1 and A2 are not products of the same gene, they share a similar domain structure: an N-terminal domain, and a glycine-rich C-terminal domain of about 10,000–12,000 daltons. The division between the N-terminal and C-terminal domain of A1 can be taken to

---

4 A. Kumar, K. R. Williams, and W. Szer, unpublished data.
occur after glutamic acid 194 which is the last amino acid that is present in a region of internal sequence homology in UP1 (29). Hence, when residues 3–93 and 94–194 in UP1 are aligned, 32% of the amino acids in these two regions are identical and an additional 29% of these changes that are seen could be accomplished by single-base changes (Fig. 6). Preliminary results of oligonucleotide-UP1 covalent cross-linking and limited proteolysis studies suggest that residues 3–93 and 94–194 represent two globular domains that can independently bind to single-stranded nucleic acids. This postulated domain structure (30) appears to be exactly analogous to that of the high mobility group protein E where limited proteolysis has been used to release two fragments corresponding to the approximately 92-amino acid residue repeat found in this protein; each of the resulting high mobility group protein E fragments is a water-soluble, ssDNA-binding protein with no apparent tendency to interact with the other (31). The apparent existence of multiple nucleic acid-binding sites in a number of eukaryotic ssDNA- and ssRNA-binding proteins suggests that this structural feature may prove to be a widely shared property of this class of proteins (30). Whereas it seems certain that the N-terminal domain of A1 and A2 is involved in interactions with RNA, the role of the C-terminal domain is not clear, but it seems to exert an effect on the nucleic acid binding properties of the protein (see below). This domain has no obvious amino acid sequence homology with the two internal repeats in the N-terminal domain.

The N-terminal domain of A1 is remarkably resistant to degradation by trypsin (Fig. 4), and this is probably true for A2 as well since limited trypsin digestion of a mixture of native A1 and A2 (Fig. 1, fractions 76–79) yields two large fragments of Mr, ~21,000 and ~22,000 (not shown). It was observed by others that treatment of 40 S monoparticles with trypsin yields a set of polypeptides that appear to be uniformly reduced in size by 11,000–13,000 daltons with respect to core proteins; these polypeptides react in Western blots with monoclonal antibodies against core proteins (5) and with antibodies against UP1 (18). The results of limited trypsin digestion of purified B2 and C1 (Fig. 8) are in line with these observations and suggest that most, perhaps all, core proteins contain a peculiarly trypsin-sensitive bond, analogous to that at Arg186 in A1 (Fig. 6), and it is the cleavage of this bond that could give rise to the set of polypeptides of Mr, between 22,000 and 28,000 regarded as ssDNA-binding proteins. It is possible, but certainly not proven, that ssDNA-binding proteins are artifacts that arise by proteolytic degradation of core proteins during cell fractionation (20, 21). Such proteolytically degraded forms of proteins that arise during isolation and retain certain in vitro activities have been reported, e.g., the calf thymus terminal deoxynucleotidyltransferase (32) and the eukaryotic fatty-acid synthetase (33). It is also possible that ssDNA-binding proteins do have an autonomous function in the cell and arise by controlled in vitro cleavage of core proteins or by differential splicing of the corresponding primary transcripts. The number of copies of a UP1-type protein has been estimated at 0.8–1.0 × 10^6 per cell (17, 19); the number of copies of a major core protein is probably about four times greater (34). In view of the results reported here, as well as (i) the immunological investigations by Pandolf et al. (20), and (ii) the lower Mr, of the mouse myeloma UP1-type protein isolated biochemically as compared to the Mr, of the species detected immunologically in crude extracts (21), the in vitro role of UP1-type ssDNA-binding proteins and the mechanism by which they are formed will have to be re-examined.

The purification of the A1 protein under nondenaturing conditions made possible an examination of some of its nucleic acid binding properties. Comparing A1 and UP1 in this respect reveals that the presence of the glycine-rich C-terminal domain modifies considerably the binding properties of the protein. A1 binds preferentially to RNA, whereas UP1 binds somewhat better to ssDNA than to RNA (28). The binding of UP1 disrupts base stacking interactions, a typical feature of RNA and DNA helix-stabilizing proteins (35), whereas the binding of A1 has no effect on the conformation of the polynucleotide ligand as judged from UV and CD spectra. Most probably related to its unwinding activity is the increased resistance of UP1-polynucleotide complexes to salt-induced dissociation as compared to A1 complexes which presumably involve primarily ionic bonds. Interactions of some aromatic amino acids with polynucleotides, in addition to ionic bonds, appear to be part of a general mechanism of the binding of helix-stabilizing proteins (36) and render their complexes relatively salt-resistant. The sensitivity of A1-polynucleotide complexes to salt is in line with the results of an earlier study which showed that A1 is among the first core proteins to dissociate when native 40 S monoparticles are sedimented in a sucrose gradient containing 0.2–0.3 M NaCl (4).

The modulatory effect of the C-terminal region of the protein on binding observed here in the A1/UP1 system suggests that this region is similar to analogous functional domains in the high mobility group protein 1 (31) and in phage T4 gene 32 (37) and E. coli SSB (38) single-stranded DNA-binding proteins. In all three cases, the proteolytic removal of a C-terminal fragment results in an increased helix-stabilizing ability of the truncated protein. In addition to limiting the unwinding activity, the C-terminal domain appears to be involved in functional protein/protein interactions, e.g. gene 32 protein and possibly also E. coli SSB protein, with components of the in vitro replication complex, and high mobility group protein 1 with histones. That the C-terminal domains of the T4 gene 32 protein and A1 hnRNP protein might be functionally similar is further suggested by the observation that both proteins contain an unusual serine-rich region within these postulated domains; gene 32 protein contains 8 serines between residues 280 and 288 (39), whereas A1 contains 6 serines extending from residue 308 to 313 (Fig. 7). Whereas it remains to be established whether the glycine-rich C-terminal domain of A1 is essential for interactions with other proteins forming the hnRNP particle, it is intriguing to find glycine-rich clusters in several nucleolar proteins (40, 41). One of these, the 110,000-dalton C23 protein is thought to be a component of preribosomal particles.

Acknowledgments—We are grateful to S. Wilson and his colleagues and to G. Reek and D. Teller for communicating their work prior to publication. We especially thank M. LoPresti, K. Stone, and N. Fernando for their excellent technical assistance. K. R. W. thanks W. Konigsberg for his continued interest in these studies.

REFERENCES


B. M. Merrill and K. R. Williams, unpublished observations.
**Supplemental Material to**

**PURIFICATION AND DOMAIN STRUCTURE OF CORE hNRNP PROTEINS A1 AND A2, AND THEIR RELATIONSHIP TO ssDNA BINDING PROTEINS**

Mallordi Kumar, Kenneth R. Williams and Michael S. Derti

**EXPERIMENTAL PROCEDURES**

**Materials**

Pituitary nucleotides were from Miles and collagenase HI mRNA was prepared as described (15). Specific radiolabeling was 14,000 cpm. Tris and DTT were from a gift from R. F. Duestach.

**Cell growth and fractionation**

Cell seats were grown in suspension culture in minimal essential medium (Dinh's modified, Gibco supplemented with 5% fetal calf serum and kept at a density between 3 x 10^6 and 6 x 10^6 cells per ml. Nuclei were obtained as described (1) and extracted by gentle stirring for four periods of 20 min each 22 x 3 volumes of a new buffer (50 mM Tris-HCl, pH 7.5, 0.0 (two extractions) or pH 8.0 (two subsequent extractions). 1.0 M NaCl, 1.5 mM MgCl, 0.1 mM DTT and 0.5 mM phenylmethylsulfonyl fluoride). The nuclear extracts were dialyzed against Tris-HCl, pH 7.5, and pelleted by ultracentrifugation as described in Materials. This suspension was dialyzed overnight against high-salt buffer, clarified by brief centrifugation at 22,000 rpm, and fractionated on a 5% to 20% sucrose gradient. The fractions containing 14,000 cpm. Tris and DTT were from a gift from R. F. Duestach.

**Nuclear protein studies**

Nucleolar proteins were precipitated with 0.2 ml of 20% trichloroacetic acid and then digested with trypsin at a protein/enzyme (w/w) ratio of 25 for 24 hr at 37°C. The resulting peptides were separated on a 5-20% linear gradient of 5.0 M sucrose, pH 7.4, column (14.5 x 62 cm) which was equilibrated with 0.25 M trichloroacetic acid at a flow rate of 0.7 ml/hr. Peptides were then eluted by increasing the concentration of acetic acid (0.5% to 100%) at a flow rate of 1 ml/min in a Beckman model 120 instrument. The fractions containing 14,000 cpm. Tris and DTT were from a gift from R. F. Duestach.

**Protein fractionation**

Nucleolar proteins were precipitated with 0.2 ml of 20% trichloroacetic acid and then digested with trypsin at a protein/enzyme (w/w) ratio of 25 for 24 hr at 37°C. The resulting peptides were separated on a 5-20% linear gradient of 5.0 M sucrose, pH 7.4, column (14.5 x 62 cm) which was equilibrated with 0.25 M trichloroacetic acid at a flow rate of 0.7 ml/hr. Peptides were then eluted by increasing the concentration of acetic acid (0.5% to 100%) at a flow rate of 1 ml/min in a Beckman model 120 instrument. The fractions containing 14,000 cpm. Tris and DTT were from a gift from R. F. Duestach.

**Filter binding assays**

The standard binding buffer for filter binding assays was 5.0 mM Tris-HCl, pH 7.6, 60 mM NaCl, and 40% glycerol, which contained 0.5 mM dithiothreitol (DTT) and 0.1% sodium dodecyl sulfate (SDS). The samples were placed on ice at room temperature, filtered through Millipore filters, washed, dried and coated with a solution of 10% dextran sulfate and 0.1% bovine serum albumin in 10 mM Tris-HCl, pH 7.4, and 0.1% SDS. After washing, the assays were used to coat membranes of complex formation. In view of the polypeptide size of most of the protein brands, their concentrations are expressed to monomeric nucleotides in legend to Tables and Figure 9.

**RESULTS**

**Purification of proteins A1, A2, B2 and C2**

Table 1 summarizes the purification of individual core hNRNP proteins under non-denaturing conditions. The results are expressed in terms of the major species, 0.3, 0.0 and 0.1% of the total counts precipitated by the various treatments. It is noteworthy that the precipitation of proteins A1, A2, B2 and C2 is obtained by a single step procedure. The data presented in this study suggest that the species of interest in the absence of DTT and in the presence of DTT, with a molecular weight of approximately 50,000 daltons, are not significantly altered in the presence of DTT.
Core hnRNP Proteins and ssDNA-binding Proteins

11273

FIG. 2. Purification of protein A1 by gel permeation chromatography in the absence of SH-reagents. a) Sephacryl S-300 column (1.6 x 120 cm) of the nuclear extract (8.0 mg/ml protein, 1.0 ml) previously dialyzed into and developed with the same buffer as in Fig. 2 except for the absence of DT. Elution pattern at 260, 320, 400, and 520 nm. A 5-fold reduced scale is shown for the peak in fractions 6-35. b) Electrophoresis of column eluates as in Fig. 1. Lane M, molecular weight markers (kDa): 30, 72, 42, 25, 16, and 14. c) Electrophoresis of column fractions no. 113-133 previously dialyzed for 18 h in a buffer containing 10 mM Tris-HCl, pH 7.4, 250 mM NaCl and 10 mM polyacrylamide (0.1 M urea per lane). The purified protein was concentrated to 0.5-1.5 mg/ml by lyophilization, clarified by low-speed centrifugation and dialyzed overnight against a buffer containing 10 mM Tris-HCl, pH 7.4, 250 mM NaCl, 1.0 mM OTT, 0.1 mM MEUF and 27% glycerol. Final yield was 0.8 mg protein. Attempts to dissociate the high molecular-weight complex containing A1 (Fig. 2b, lane 12) with SH-reagents and/or high salt in the absence of urea were unsuccessful. Some preparations of A1 used in this study were obtained by elution from SDS-polyacrylamide gels (Fig. 3).

While at IL-13, who isolated proteins A1 by electrophoresis from SDS-polyacrylamide gels, observed the somewhat erratic appearance of the A1 doublet and, after analyzing the entire anti-cytochrome c antibody-generated polypeptide map of the two bands concluded that they were identical. They also showed that A1 migrates as a single band if 8.0 mM urea is included in the gel. In addition to SDS-PAGE, we do not observe the A1 doublet when some proteins are fractionated in the presence of SH-reagents (Fig. 1b), and we find that the doublet virtually disappears, i.e., a single band is seen in the standard SDS/polyacrylamide gel (Fig. 2c) if the material purified in the absence of SH-reagents is eluted overnight against a buffer containing a high concentration of DT (10 mM Tris-HCl, pH 7.4, 250 mM NaCl, 10 mM DTI). The appearance of the doublet could be the result of cystine formation. Further analysis and sequencing of A1 confirms that the material purified as described is not a mixture of different polypeptides. The yield of nearly homogeneous A1 varies from 0.25 to 0.6 mg from 16 mg protein in the nuclear extract. The proteins can be kept frozen for at least 4-5 months at -20°C; it gradually loses binding activity on thawing and freezing.

For the purification of protein A2, fractions of the Sephacryl S-300 column enriched in A2 (Fig. 2b, fractions 6 to 15, total protein 2.4 mg) were dialyzed against a buffer containing 24 mM sodium acetate, pH 5.5, 1.0 mM DTI, 0.1 mM MEUF and 0.2 M urea for 16 h with 3 changes, and loaded on a Di-O-cellulose column (0.5 x 10 cm) equilibrated in the same buffer. The column was developed with 40 ml at a linear gradient of 0-200 mM NaCl in the same buffer. Fractions of 0.2 ml were collected and ali-quots (about 10 mg protein) were analyzed in Fig. 1. Fractions 14 and 15 (Fig. 2a), eluting as about 170 mM NaCl were pooled, concentrated, and chromatographed on a Sephadex G-100 column (0.27 x 90 cm) equilibrated in the same buffer containing 10 mM Tris-HCl, pH 7.4, 250 mM NaCl, 1.0 mM OTT, 0.1 mM MEUF and 27% glycerol. Final yield was 0.4 mg protein. Attempts to dissociate the high molecular-weight complex containing A2 (Fig. 2b, lane 12) with SH-reagents and/or high salt in the absence of urea were unsuccessful. Some preparations of A2 used in this study were obtained by elution from SDS-polyacrylamide gels (Fig. 3).

For the purification of protein A2, fractions of the Sephacryl S-300 column enriched in A2 (Fig. 2b, lanes 11 to 15, total protein 3.2 mg) were pooled, dialyzed for 16 h against a buffer containing 10 mM Tris-HCl, pH 7.4, 250 mM ammonium bicarbonate, concentrated to 1.0 ml by centrifugation, clarified by low-speed centrifugation and dialyzed for 16 h against 10 mM Tris-HCl, pH 7.4, 500 mM NaCl, 1.0 mM OTT, 0.1 mM EDTA and 0.1 mM MEUF. The sample was chromatographed on a Sephadex G-100 column equilibrated in the same buffer and analyzed as described in Fig. 1. Fractions containing about 85% pure A2 (0.6 mg) were pooled and concentrated on Amicon 10 filters (Fig. 4, lane 13).