Isolation and partial characterization of U1–U6 small RNAs from Bombyx mori

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 1985

Impact Factor: 3.02 · DOI: 10.1021/bi00322a017 · Source: PubMed

4 AUTHORS, INCLUDING:

David S Adams
Worcester Polytechnic Institute
40 PUBLICATIONS **3,094** CITATIONS

Reprinted from Biochemistry, 1985, *24*, 117.

Isolation and Partial Characterization of U1–U6 Small RNAs from *Bombyx mori*†

David S. Adams,*,‡ Rene J. Herrera,‡ Reinhard Luhrmann,§ and Paul M. Lizardi‡

Laboratory of Cell Biology, The Rockefeller University, New York, New York 10021, and Max-Planck-Institut für Molekulare Genetik, Otto-Warburg-Laboratorium, D-1000, Berlin 33, West Germany
Received May 30, 1984

ABSTRACT: We have used a variety of techniques to characterize the U-series small nuclear RNAs from the posterior silk gland of *Bombyx mori*. Six molecular species have been identified which correspond to the vertebrate U1–U6 RNAs by the following criteria: (a) presence of the RNAs in ribonucleoprotein particles which can be immunoprecipitated by lupus Sm antisera; (b) presence of a 2,2,7-trimethylguanosine cap, as assayed by immunoprecipitation with anti-2,2,7-trimethylguanosine IgG; (c) size, as assayed by acrylamide/urea gel electrophoresis using HeLa cell U-RNA markers; and (d) primary nucleotide sequence, as determined by chemical/enzymatic cleavage of end-labeled molecules. The high conservation of primary sequence (66–81% homology based on partial sequences) relative to the corresponding vertebrate U-RNAs has permitted unambiguous identification of each molecule. With the exception of two subspecies of U3 RNA, the U-snRNAs of *Bombyx* exhibit a striking conservation of secondary structure relative to the proposed structures of the U-RNAs of vertebrates. This conservation is best exemplified by several compensatory base alterations that result in the maintenance of hairpin structures. These are particularly evident in U1 and U5 RNAs. *Bombyx* U3 is interesting in that two subspecies (of a total of four that were sequenced) diverge considerably in sequence (and presumably in structure) relative to the U3 RNA of vertebrates. The most abundant U-RNAs in the posterior silk gland appear to be U1 and U2, while U3–U6 are present in relatively small amounts.

In the analysis of the control of gene expression, one area of current intense interest is the role played by small RNA molecules (sRNAs).[1] These molecules are low molecular weight, nonribosomal, nontransfer RNAs that have been found, variously represented, in every eukaryotic organism examined to date. Different types of sRNAs have been pro-

† This work was supported by Grant GM 28986 from the National Institute of General Medical Sciences awarded to P.M.L.

* Address correspondence to this author at the Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA 01609. At The Rockefeller University, he was a Postdoctoral Research Fellow of the American Cancer Society (Grant PF-1985) and of the U.S. Public Health Service (Grant GM 08382).

‡ Laboratory of Cell Biology, The Rockefeller University.

§ Max-Planck-Institut für Molekulare Genetik.

[1] Abbreviations: snRNP, small nuclear ribonucleoprotein; sRNA, small RNA; snRNA, small nuclear RNA; $m_2^2m^7G$, N^2,N^2,7-trimethylguanosine; EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; PBS, 10 mM Na_2HPO_4 and 0.15 M NaCl, pH 7.5; NET, 150 mM NaCl, 5 mM Na_2EDTA, and 50 mM Tris-HCl, pH 7.5; SSC, 15 mM sodium citrate and 0.15 M NaCl, pH 7.5; kb, kilobase; TBE, 90 mM Tris base, 90 mM borate, and 2 mM Na_2EDTA; RNP, ribonucleoprotein; PSG, posterior silk gland; pCp, cytidine bisphosphate.

0006-2960/85/0424-0117$01.50/0

posed to function in such varied processes as transcriptional stimulation (Kanehisa et al., 1974; Goldstein, 1976; Ringuette et al., 1980; Sohn et al., 1983), RNA splicing (Murray & Holliday, 1979; Lerner et al., 1980; Rogers & Wall, 1980), translational control (Bester et al., 1975; Thimmappaya et al., 1982), and translocation of secretory proteins (Walter & Blobel, 1982). Other postulated functions are initiation of DNA replication (Jelinek et al., 1980), transcriptional termination (Galli et al., 1983), formation of chromatin tertiary structure (Pederson & Bhorjee, 1979), and mRNA transport (Harada & Kato, 1980).

The six U-series sRNAs (U-RNAs, U1–U6) constitute the most highly characterized set of sRNAs [see Busch et al. (1982) for a review]. Vertebrate U-RNAs range in size from 105 nucleotides (for U6) to 218 nucleotides (for U3) and exist associated with protein in RNP particles (snRNPs). U-RNAs contain several types of internal base modifications, are rich in uridylic acid, contain no poly(adenylic acid), and, with the exception of U6, have a trimethylated 5′-cap structure. The complete nucleotide sequence of all six U-RNAs is known for several species of vertebrates, including humans. No evidence currently exists for developmental or tissue specificities of any of the U molecules (Hellung-Larsen & Frederiksen, 1977; Zeller et al., 1983; Forbes et al., 1983).

In higher eukaryotes, U1 RNA is the most abundant of the U-RNAs. This molecule is of great interest because of growing body of circumstantial evidence implicates U1 RNP as a major component of the mRNA splicing machinery of the cell nucleus. Several years ago, U1 RNA was proposed to play a role in the recognition and alignment of pre-mRNA splice junctions (Lerner et al., 1980; Rogers & Wall, 1980). Soon thereafter, Yang et al. (1981) were able to show that lupus antibodies, which preferentially precipitate U1 RNP, also will inhibit the splicing of adenovirus type 2 early transcripts in isolated nuclei. In more recent experiments, Padgett et al. (1983) have used anti-U1 RNP antibodies and anti-Sm monoclonal antibodies to demonstrate that the U1 ribonucleoprotein is essential for in vitro splicing of adenovirus late mRNA. Direct binding of U1 RNP to a 5′-splice junction in an in vitro system has been demonstrated by Mount et al. (1983). It is believed that this binding interaction is mediated in part by the sequence ACUUACCUGG which occurs at the 5′ end of U1 RNA. It is known from structural probing studies that this terminal domain of the U1 molecule is accessible to base pairing in deproteinized U1 RNA (Lazar et al., 1982), as well as in U1 RNP (Rinke et al., 1984). It is striking that the sequence of 10 nucleotides near the 5′ end of U1 RNA is conserved in all U1 RNA molecules studied to date. This finding underscores the importance of studying in detail the evolution of primary sequence as well as secondary structure of these molecules, since such studies may reveal important functional domains.

Relatively little is known about the U-RNAs of invertebrate species. Recent work indicates that U-RNAs are probably found in all eukaryotes but are 10–100-fold less abundant (compared to ribosomal RNAs) in the more primitive eukaryotic phyla. This low abundance may imply that the U-RNAs do not function in any capacity directly related to cell size, generation time, or cell architecture (Wise & Weiner, 1981). So far, only a few U-RNAs from invertebrates have been characterized by sequencing. These include *Dictyostelium* (Wise & Weiner, 1980; Takeishi & Kaneda, 1981), *Drosophila* (Mount & Steitz, 1981; Wooley et al., 1982), *Tetrahymena* (Branlant et al., 1983), Dinoflagellates (Reddy et al., 1983), pea plant (Krol et al., 1983), and yeast (Wise et al., 1983). Most of these studies identified the molecules as U-RNAs on the basis of size (although none exactly comigrated with their vertebrate counterparts), precipitability with lupus antibody (Lerner & Steitz, 1979), precipitability with anti-$m_2^2m^7G$ antibody (Bringmann et al., 1983), or by primary sequence homology. The lupus antigen(s), the 5′-$m_2^2m^7G$ cap, and the internal base modifications are probably conserved throughout all eukaryotic organisms, and they are found on the U-RNAs of the primitive Dinoflagellates (Reddy et al., 1983).

Our laboratories have been interested in the characterization of small nuclear RNAs in the silkworm *Bombyx mori*, an organism well suited for the study of developmentally regulated processes. In this paper, we report the identification, isolation, and partial characterization of six *Bombyx* U-RNAs. As with vertebrate U-RNAs, these molecules contain a $m_2^2m^7G$ cap and exist associated with protein (snRNPs). Although the *Bombyx* molecules differ slightly in length from their vertebrate counterparts, at the primary sequence level they exhibit a relatively high degree of sequence conservation (66–81% homology) with vertebrate U-RNAs. Species U1 and U2 are most abundant in the highly specialized posterior silk gland.

Materials and Methods

Biological Material. Bombyx mori eggs (non-inbred strain R, originally derived from a Gunka X Hoshun cross) were purchased from Syunsure Shibata Gunze LTD., Japan, and were stored at 9 °C for up to 3 months before use. Methods for raising larvae on an artificial diet have been described (Suzuki & Brown, 1972). Bm-5 cells, originally established in 1967 from ovarian tissue (Grace, 1967), were obtained from Dr. James Vaughn from the Insect Pathology Laboratory of the U.S. Department of Agriculture, Beltsville, MD. Cells were maintained at 28 °C in TC-100 medium (Gardiner & Stockdale, 1975) supplemented to 10% with fetal calf serum. HeLa were kindly supplied by Dr. Joseph Nevins of The Rockefeller University.

Culture and Larval Labeling. Larvae of the late fifth instar (day 5) were labeled with [^{32}P]orthophosphate (carrier free, New England Nuclear) by direct addition of the label to the diet. In a typical experiment, 2.0 mCi of ^{32}P (in 200 μL of 0.02 M HCl) was administered per larva and, after a 24-h incubation, resulted in RNA with a specific activity of (2–5) $\times 10^4$ cpm/μg. Larvae were labeled with [5-^{3}H]uridine (sterile aqueous, NEN) by intracoelomic injection as described (Suzuki & Brown, 1972). Typically, 2.0 mCi of [^{3}H]uridine in 100 μL of 1 × SSC, pH 7.5, was administered per larva and, after 24 h of incubation, gave RNA with a specific activity of (1.0–3.5) $\times 10^4$ cpm/μg. Bm-5 and HeLa cells were labeled with [5-^{3}H]uridine (sterile aqueous, NEN) by direct addition of the label to the culture medium to 10 μCi/mL. RNA of specific activity 1×10^4 (Bm-5) or 8×10^4 (HeLa) cpm/μg was obtained after 24 h of incubation.

Cell Fractionation and RNA Extraction. Except for immunoprecipitation experiments (see below), HeLa and Bm-5 nuclear and cytoplasmic fractions were prepared as described (Penman, 1966). RNA was prepared from nuclei by extraction with hot phenol (Penman, 1966) and from posterior silk glands or cytoplasmic fractions by the proteinase K/sodium perchlorate procedure (Lizardi & Engelberg, 1979).

Immune Sera. Sera from patients with systemic lupus erythematosus (types anti-Sm and anti-RNP) were kindly provided by Dr. Robert Lahita at The Rockefeller University. IgG was purified by ammonium sulfate fractionation followed by DEAE chromatography (Garvey et al., 1977), was dissolved in PBS, pH 8.5, to 15 mg/mL, and was stored at −20 °C for

up to 2 years prior to use. Anti-$m_2^2m^7G$ antibodies were induced and isolated as described (Luhrmann et al., 1982).

Immunoprecipitations. These assays were conducted essentially as described previously for lupus (Lerner & Steitz, 1979) or anti-cap IgG (Bringmann et al., 1983). Cultures labeled in vivo were used for quantitative (mass) comparisons of precipitated RNAs, whereas RNA isolated from unlabeled cultures was end labeled in vitro for sequence analysis.

For HeLa and Bm-5 lupus precipitations, nuclear fractions were prepared from 1×10^8 cells. For posterior silk gland precipitations, homogenates were prepared from two gland pairs by homogenization into 2.0 mL of lysis buffer (10 mM Tris-HCl, pH 8.5, 100 mM NaCl, and 1 mM $MgCl_2$) using a loose (type A) glass Dounce homogenizer. These fractions were sonicated (Lerner & Steitz, 1979), and the sonicate was cleared by a 1-min centrifugation in an Eppendorf microfuge. Aliquots representing 10^7 Bm-5 or HeLa nuclei, or two gland pairs, were incubated for 15 min at 0 °C with 200 μg of purified IgG. Immune complexes were collected by adding the pellet representing 100 μL of a 10% suspension of freshly washed Pansorbin (Calbiochem) (Kessler, 1975). After the incubation was continued for 15 min at 0 °C, the Pansorbin-bound complexes were washed 5 times in NET, pH 7.5, in an Eppendorf microfuge using 15-s centrifugations. RNAs were purified by extraction with phenol and were analyzed by electrophoresis at 50 °C in 40 cm long, 0.8 mm thick gels which contained 10% acrylamide [27:1 acrylamide:bis-(acrylamide)], 8 M urea, and 45 mM TBE, pH 8.3 (Donis-Keller et al., 1977). Autoradiography or fluorography was conducted as described (Bonner & Laskey, 1974).

In some experiments, unlabeled RNA fractions were subjected to chromatography on Bio-Gel A (1.5 m) to remove transfer RNA which prevents efficient end labeling of other sRNAs. Aliquots (0.4 μg) were end labeled with ^{32}P by using RNA ligase (P-L Biochemicals) and [^{32}P]pCp (NEN).

For anti-cap precipitations, a 1.0-mg aliquot of purified RNA (total nuclear for Bm-5 or HeLa and total gland for PSG) was mixed with 150 μg of anti-cap IgG in a total volume of 200 μL of a solution containing PBS, pH 7.5. After a 1-h incubation on ice and a clarifying spin of 15 s in a microfuge, Pansorbin was used to collect the immune complexes as indicated above. Pansorbin pellets were washed 5 times with either a low-stringency wash (NET, pH 7.5, and 0.05% Triton X-100) or a high-stringency wash (0.4 M NaCl, 0.4 M sodium acetate, pH 5.0, 5 mM Na_2EDTA, 50 mM Tris-HCl, and 0.05% Triton X-100). The high-stringency wash has been shown to decrease nonspecific binding (Bringmann et al., 1983). RNAs were analyzed as indicated above.

RNA Sequencing. Unlabeled immunoprecipitated RNA or total RNA fractionated on a TSK-G-3000 column (LKB) to enrich for various sRNAs was 3′ end labeled to high specific activity (10^7–10^8 cpm/μg) using [^{32}P]pCp (NEN) and RNA ligase (P-L Biochemicals) (England & Uhlenbeck, 1978). RNAs were fractionated on partially denaturing gels (as described above), and individual bands were located by wet autoradiography and excised by high-salt extraction (Gilbert & Maxam, 1973). Individual molecules were sequenced by using both the direct chemical method of Peattie (1979) and the enzymatic method [see Vournakis et al. (1981) for a review].

Results and Discussion

Immunoprecipitation of Bombyx snRNPs with Human Lupus Antibody. Antibodies isolated from the sera of patients with the autoimmune disease systemic lupus erythematosus have previously been shown to react with antigens located on

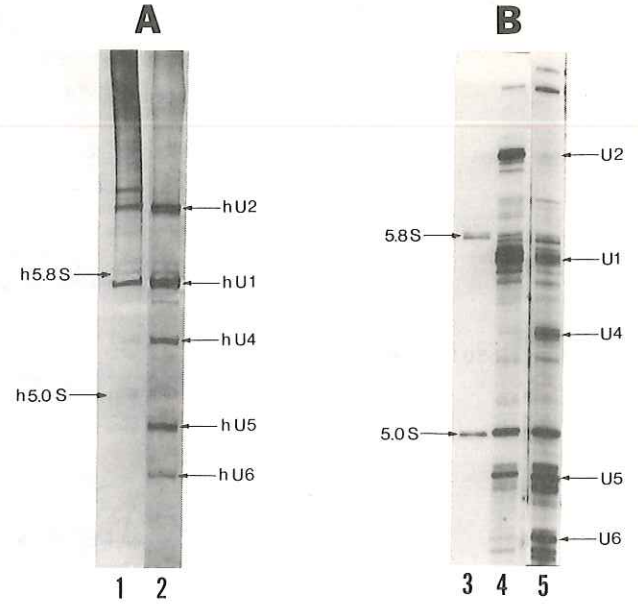

FIGURE 1: Immunoprecipitation of RNA with lupus IgG. Panel A shows the specificity of our sera for HeLa RNP. Purified anti-Sm IgG (lane 2) was used to precipitate ^3H-labeled (24 h) HeLa nuclear RNP. Immune complexes were collected by using Pansorbin, and the purified RNAs were analyzed by electrophoresis on acrylamide/urea/TBE gels run at 50 °C. Lane 1 represents 100 000 cpm of total HeLa nuclear RNA. Lane 2 contains 10 000 cpm. Exposure for fluorography was for 1 week at −70 °C with one intensifying screen. Panel B shows the electrophoretic analysis of lupus-precipitated *Bombyx* RNA. Precipitation was performed with unlabeled lysates, and purified RNAs were end labeled with ^{32}P to allow rapid sequence analysis. Lanes 4 and 5 compare the RNAs precipitated by anti-Sm IgG from total posterior silk gland lysates (lane 4) (10 000 cpm) and Bm-5 nuclear lysates (lane 5) (10 000 cpm). Lane 3 is a marker representing 2000 cpm of total PSG RNA labeled with ^{32}P in vivo. Exposure for autoradiography was for 2 days at −70 °C with one screen.

U-RNPs from a variety of organisms, including man, mouse, fall armyworm (Lerner & Steitz, 1979; Lerner et al., 1980), *Drosophila* (Mount & Steitz, 1981; Wooley et al., 1982), and Dinoflagellates (Reddy et al., 1983). Current evidence implicates a 26 000-dalton protein as the conserved and ubiquitous antigen recognized by anti-Sm IgG (present on all U-RNAs except U3) (Lerner et al., 1981; Wooley et al., 1982; Wieben et al., 1983). A 32 000-dalton protein may be the U1-specific antigen recognized by anti-RNP IgG (Wieben et al., 1983).

Since insect systems contain a relatively low concentration of small RNPs, we screened a variety of lupus sera in an effort to identify high-titer IgG fractions. The specificity of IgG preparations was tested by using HeLa cell nuclear extracts (Lerner & Steitz, 1979). The identity of the precipitated HeLa RNAs was determined by reference to published profiles (Lerner & Steitz, 1979; Lerner et al., 1980) and by size estimation relative to RNA standards. An anti-Sm serum was chosen for its high titer and its ability to precipitate all HeLa U-RNAs (except U3) (Figure 1A).

Two sources of *B. mori* material were used in our initial trials with lupus antibodies: nuclear extracts of Bm-5 cells, a permanent cell line, and posterior silk gland (PSG), a highly specialized larval tissue. The ramified morphology of PSG nuclei precludes cell fractionation; thus, in this latter case, immunoprecipitations were performed on whole tissue lysates. Immunoprecipitated RNAs were terminally labeled in vitro to permit subsequent identification by partial sequencing. Figure 1B shows the pattern of U-RNAs obtained from *B. mori* material using anti-Sm IgG. The high-resolution (80 cm) gel coupled with partial sequence analysis permitted the

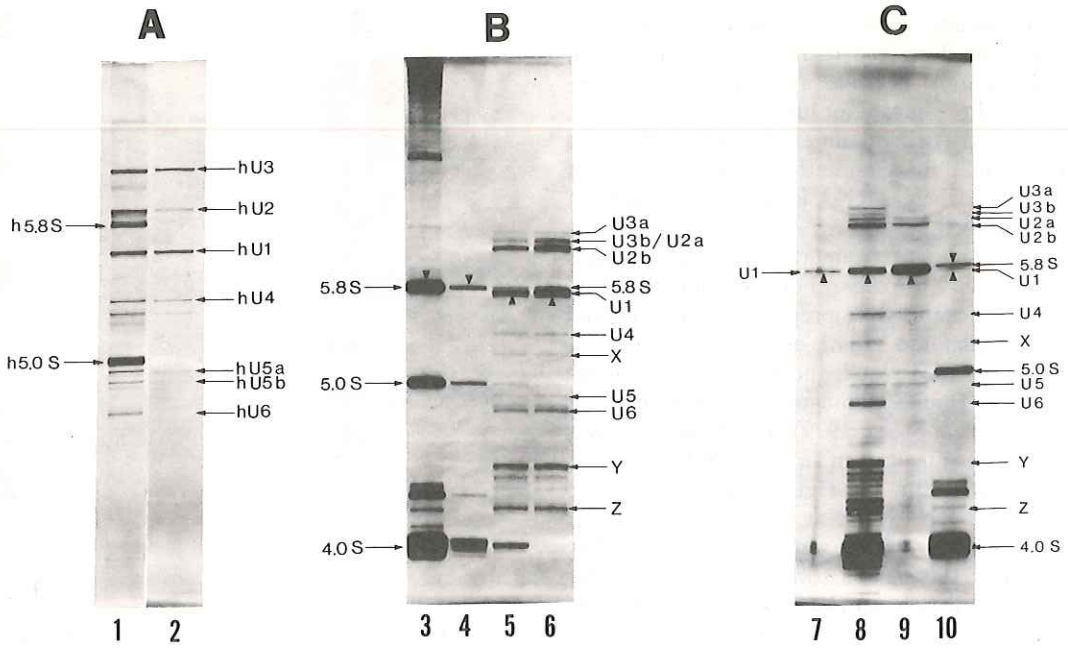

FIGURE 2: Immunoprecipitation of RNA with anti-cap IgG. Panel A shows the specificity of the IgG for HeLa RNA. ^{3}H-Labeled (24 h) HeLa nuclear RNA was used for immunoprecipitation. Electrophoresis of anti-cap-precipitated RNA is shown in lane 2 (1000 cpm), and total nuclear RNA marker is shown in lane 1 (80 000 cpm). Fluorography was for 20 days at −70 °C with one screen. Panel B shows the electrophoretic analysis of anti-cap-precipitated *Bombyx* RNA. The gland RNA used for precipitation was labeled in vivo with ^{32}P for 24 h. Lanes 3 and 4 show markers of 80 000 and 8000 cpm of total posterior silk gland RNA, respectively. Lanes 5 and 6 represent 1000 cpm of anti-cap-precipitated gland RNA where the Pansorbin-bound RNA was washed with a low-stringency wash (lane 5) or a high-stringency wash (lane 6). Autoradiography was for 1 day at −70 °C with one screen. Panel C represents a comparison of lupus and cap-precipitated *Bombyx* RNA. The anti-Sm-precipitated RNAs (lanes 7 and 9) were chromatographed on Bio-Gel A (1.5 m) to remove transfer RNA which prevented efficient postlabeling of the U-RNAs. 200 (lane 7) and 2000 cpm (lane 9) of anti-Sm-precipitated PSG RNA were layered on the gel. Lane 8 represents a 2000 cpm portion of postlabeled anti-cap-precipitated total PSG RNA which was used for sequence analysis. Lane 10 shows 5000 cpm of postlabeled total PSG gland RNA as marker. Autoradiography was for 1 day at −70 °C with one intensifying screen.

unequivocal identification of U1, U2, U4, U5, and U6 RNAs in both PSG (lane 4) and Bm-5 (lane 5). Some of the bands occur as doublets or triplets, which may reflect alternative conformations or sequence microheterogeneity, since at the sequence level each group shares the same 3′ terminus. 5.0S ribosomal RNA occurred as a contaminant in these preparations due to nonspecific binding to the Pansorbin (see the next section on anti-cap IgG precipitations). High-stringency washes significantly reduced this contamination but resulted in differential losses of U-RNAs; therefore, low-stringency washes were used in experiments designed to analyze the relative abundance of each U-RNA species. Relative to the Bm-5 cells, PSG material contains relatively little U4 and U6 RNA. Conversely, U2 and U1 are recovered in lower amounts from Bm-5 cells, but we have not yet excluded partial degradation as a possible reason for this difference (one should bear in mind that Bm-5 cells were subjected to nuclear fractionation, while PSG was not). As is the case in most organisms so far studied, little, if any, U3 RNA is recovered by using anti-Sm IgG.

From this point on, further analysis of *B. mori* U-RNAs was carried out by using PSG material, which gave the highest yield of U1 and U2 RNA. However, we needed a method which would allow us to isolate U3 RNA, given that the lupus antibodies did not precipitate this molecule.

Immunoprecipitation of U-RNAs with Anti-Cap IgG. With the exception of U6, all U-RNAs analyzed to date from a variety of eukaryotic organisms contain a 5′-$m_2^2m^7G$ cap. This cap contrasts with the 7-methylguanosine (m_7G) cap found in mRNA and provides a strong criterion for U-RNA identification. Antibodies induced against $m_2^2m^7G$ have recently been isolated and characterized (Luhrmann et al., 1982; Smith & Eliceiri, 1983). These antibodies have been shown to specifically precipitate all U-RNAs (except U6 which is not capped) from mouse (Bringmann et al., 1983) and human KB cells (Smith & Eliceiri, 1983) and to precipitate U1, U2, and U5 from pea nuclei (Krol et al., 1983).

We utilized anti-cap IgG to isolate *B. mori* U-RNAs. The analysis of precipitated RNA from the PSG was carried out by using material labeled homogeneously (^{32}P in vivo) for quantitation of U-RNAs (Figure 2, panel B), as well as with end-labeled RNA (^{32}P in vitro) for sequencing analysis (Figure 2, panel C). The pattern of mass-labeled, immunoprecipitated sRNAs in lanes 5 and 6 illustrates the effects of low- and high-stringency washes of the Pansorbin/immune complexes. High-stringency washes (high-salt NET) have been shown to decrease nonspecific binding of RNA to Pansorbin (Bringmann et al. 1983). Low-stringency washes were sufficient for removing nonspecifically bound RNA (5S and transfer RNA) for the HeLa cell controls (panel A, lane 2). For PSG, high-stringency washes are required for the removal of 5S RNA and 4S RNA contaminants, but these washes may result in differential losses of U-RNAs; hence, the pattern of the U-RNAs isolated from PSGs by low-stringency washes (Figure 2, lane 5) is probably a more accurate representation of the relative abundancy of U-RNAs in the tissue. Examination of this pattern reveals the following: (1) U1 and U2 are by far the most abundant U-RNAs in PSG. (2) Several species of U3 RNA exist in PSG, the most abundant being U3a and U3b as shown in Figure 2B,C. High-resolution (80 cm) gels reveal a total of four U3 species (a, b, c, and d; see sections below for sequencing data). (3) PSG appears to contain a capped U6 RNA which is recovered in excellent yield after high-stringency washes. This last result is in contrast to HeLa U6, which is not recovered by anti-cap IgG immunoprecipitation (compare lanes 2 and 6 in Figure 2).

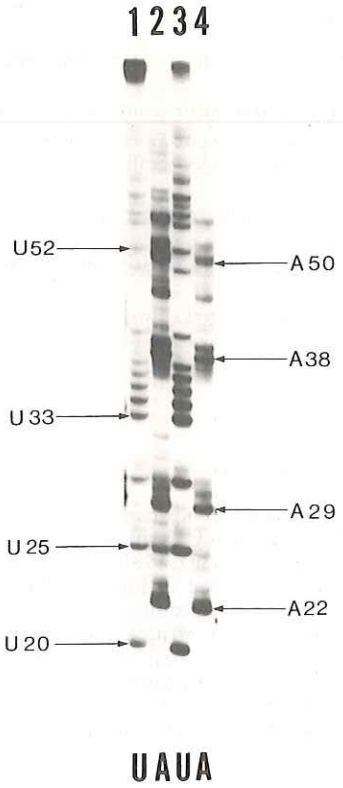

FIGURE 3: Sequence fingerprint comparison of *Bombyx* U1 RNAs. End-labeled *Bombyx* U1 RNA isolated from PSG by immunoprecipitation with anti-Sm IgG (lanes 1 and 2) or by precipitation with anti-cap IgG (lanes 3 and 4) was chemically cleaved in adenosine-specific (A > G) (lanes 2 and 4) or uridine-specific reactions (U ≫ G) (lanes 1 and 3). 5000 cpm were applied to each gel lane. Exposure for autoradiography was for 1 week at −70 °C with one screen. The numbering of each nucleotide was established by complete sequence analysis of the material in lanes 3 and 4 using procedures similar to those described in Figure 4.

Initially we considered the possibility that U6 was recovered in the immunoprecipitated fraction due to its ability to form complexes with U4 RNA (Hashimoto & Steitz, 1984; Bringmann et al., 1984). However, more recently we have carried out experiments where unlabeled sRNAs are denatured in formamide, separated in a TBE/urea/acrylamide gel, electroblotted to a membrane, and finally detected in the blot with radiolabeled anti-$m_2^2m^7G$ antibody. In these experiments (Adams et al., 1984), U6 RNA is detected directly on the membrane by antibody, implying that the molecule must contain some kind of modified guanosine structure (which may or may not be an inverted cap).

Results of the partial sequencing which allowed unequivocal identification of the RNAs are given in a later section. RNAs precipitated by both types of IgG (anti-Sm or anti-cap) are indistinguishable by mobility on gels (Figure 2, lanes 8 and 9). Note that tRNA does not contaminate the material in lanes 7 and 9 since it was removed by column chromatography to provide efficient in vitro labeling of the U-RNAs for sequencing. Transfer RNA was not removed from the anti-cap IgG-precipitated material prior to in vitro labeling since quantities sufficient for sequence analysis were easily generated. Figure 3 illustrates a comparison of uridine-specific chemicl sequencing reactions (lanes 1 and 3) and adenosine-specific reactions (lanes 2 and 4) for U1 RNA obtained by using either type of immunoprecipitation. While not exhaustive, this "sequence-fingerprint" analysis does allow a rapid comparison of two molecules in a manner analogous to an "RNase fingerprint". Here, it suggests that both immunoprecipitation methods yield identical species of U1, U2, U4, U5, and U6 RNAs. The cap-immunoprecipitated RNAs marked by the symbols Y and Z in Figure 2 have not been sequenced. Molecule X is discussed below.

Table I: *Bombyx mori* U-RNAs

vertebrate U-RNAs		*Bombyx* U-RNAs			
molecule	length (nucleotides)[a]	molecule	length (nucleotides)[b]	% of molecule sequenced	% homology with rat U-RNAs[a]
U3a	216	U3a[d]	~210	57	79
U3b	216	U3b[d]	206	60	75
U3c	216	U3c[d]	~211	85	66
		U3d[d]	201	37	68
U2	189	U2a	199	60	78
		U2b	196	60	78
U1	165	U1	163	82	73[c]
U4a	146				
U4b	145	U4	144	72	81
U4c	142				
U5a	117	U5	115	85	79
U5b	115				
U6	107	U6	109	49	79

[a] Vertebrate U-RNA lengths and primary sequences have been published [see Busch et al. (1982) for a review]. [b] Determined by comparison of electrophoretic mobilities of these molecules with the mobilities of mammalian sRNAs whose exact lengths are known from complete primary sequence analysis [see Busch et al. (1982) for a review]. [c] This *Bombyx* U1 also exhibits ~90% homology with the *Drosophila* U1 sequence previously published by Mount & Steitz (1981). [d] The *Bombyx* U3a–d nomenclature does not necessarily correspond to that of the rat U3s.

Size of Bombyx U-RNAs. The approximate lengths of the RNAs were estimated by comparison of their electrophoretic mobilities with the mobilities of mammalian sRNAs whose exact lengths are known from complete primary sequence analysis [for a review, see Busch et al. (1982)]. For extrapolation purposes, we used a plot of log mobility vs. the square root of molecular weight (Lehrach et al., 1977) (data not shown). Estimated *Bombyx* U-RNA sizes are listed in Table I. As is the case for the nonvertebrate U-RNAs analyzed to date, the *Bombyx* U-RNAs do not exactly comigrate with their vertebrate counterparts. The *Bombyx* molecules most closely resembling their vertebrate counterparts in size were U1, U4, U5, and U6, which were within a few nucleotides of the expected size. In contrast, U2 and U3 differed by 6–10 nucleotides from their vertebrate counterparts.

Sequence of B. mori U-RNAs. Table I lists the various *B. mori* U-RNAs that were partially sequenced and the extent of sequence homology with the corresponding molecules in the rat. The most easily readable and unambiguous sequencing gel patterns were obtained for several subspecies of U3 RNA, as well as for U1 and U5 RNA. The sequences of these molecules will be discussed in Figures 5–7 using published rat or insect sequences as frames of reference. Molecule X (Figure 2, lanes 5, 6, and 8) has been partially sequenced and is not homologous to any known sRNA (unpublished data).

U3 RNA presented the most complex situation since four subspecies were detectable by sequencing. These molecules are interesting because their relative abundancies are different in PSG and Bm-5 cells, a problem which is being investigated further. Figure 4 shows an example of the combined chemical/enzymatic sequencing of two sequence variants of *B. mori* U3 RNA, U3c and U3d. Lanes 1–4 and 6–9 represent the chemical technique, while lanes 5 and 10 represent guanosine-specific enzymatic cleavage. Panel A represents U3c while panel B shows U3d. As has been reported by other laboratories (Mount & Steitz, 1981), some regions of the sequencing gels

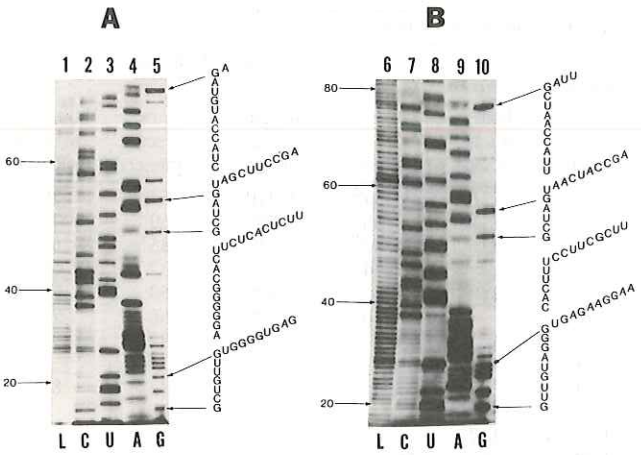

FIGURE 4: Primary sequence analysis of *Bombyx* U3 RNAs. Total PSG RNA was fractionated on a TSK-G-3000 column (LKB) to enrich for U3 RNA. This RNA was end labeled with ^{32}P and fractionated on acrylamide/urea/TBE gels to separate U3c (panel A) from U3d (panel B). RNAs were excised by extraction with high salt and subjected to chemical (lanes 1–4 and 6–9) or enzymatic (lanes 5 and 10) base-specific cleavage reactions. Randomly cleaved molecules (ladders) are shown in lanes 1 and 6. Cytosine-specific (C > U) reactions are in lanes 2 and 7, uridine-specific reactions (U ≫ C) are in lanes 3 and 8, adenosine-specific (A > G) reactions are in lanes 4 and 9, and enzymatically cleaved guanosine-specific reactions are in lanes 5 and 10. 10 000 cpm were applied to all lanes except for "ladders" which contained 50 000 cpm. Exposure for autoradiography was for 1 week at −70 °C with one screen.

```
Drosophila U1:  5' GmpppAmUmACΨΨACCUGGCGUAGA
                    o     o o oo**ooo*oooo* o*
Bombyx U1:      5' N     N N NNuuNNNuNNNNugNa

GGUUAACCGUGAUCACGAAGGCGGUUCCUCCGGAGUGAGGCUUG
*ooo**oo**o**o*****oooo****** **o***ooo******
gNNNAANNGUNAUNACGAANNNGGUUCCACCNGAGNNNGGCUUG

GCCAUUGCACCUC-GGCUGAGUUGACCUCUGCGAUUAUUCCUAA
*  **********   *** *******  ****************
G--AUUGCACCUCCAACUGUGUUGACCGUUGCGAUUAUUCCUAA

UGUGAAUAACUCGUGCGUGUAAUUUUUGGUAGCCGGGAAUGGCG
****  *******  ** ***************  *** * ***
UGUGGGUAACUCGCACGCGUAAUUUUUGGUAGCGUGGACU-GCG

UUCGGGCCGUCCCGA- 3'
* ***** ****o  o
UACGCGCUGUCCNCGN 3'
```

FIGURE 5: Primary sequence comparison for U1 RNAs of *Drosophila* and *Bombyx*. The *Drosophila* U1 sequence is that published by Steitz and co-workers (Mount & Steitz, 1981). The sequence for *Bombyx* U1 was obtained as described in Figure 4 for U3 RNAs. The alignments were chosen to maximize homology. Bases differing between the two molecules are written in the *Bombyx* sequence and have no asterisk between them. Asterisks designate base similarities, open circles designate unknown bases, and hyphens designate base deletions. The sequences are written 5′ to 3′. Limitations of chemical and enzymatic direct RNA sequencing limit the reliability of base assignments.

showed poor pyrimidine discrimination. This problem may result from residual RNA secondary structure in the base-specific cleavage reactions. The enzymatic cleavages are more dependent upon the secondary structure of the region being cleaved than are the chemical reactions. This is evidenced by the enzymatically cleaved guanosine-specific bands which are less prominent than their chemically cleaved counterparts (Figure 4, lanes 4 and 5, nucleotides 31–35). In lane 4, the guanosine residue at position 31 is obscured by the "overexposed" adenosine residue in position 30; it is easily seen in a shorter time of autoradiographic exposure (data not shown). Interestingly, the region where the above-mentioned discrepancy is observed corresponds exactly to the expected stem regions for *Bombyx* U3 (see Secondary Structure of *Bombyx* U-RNAs below).

In terms of relative homology compared to rat U-RNAs, the *B. mori* U3 RNAs are the most divergent molecules (i.e., 66% conservation for U3c). The other *B. mori* U-RNAs show sequence homologies of 73–81% when compared to the rat. This latter degree of conservation is analogous to that observed for *B. mori* 5S RNA (76% homology relative to the rat; Troutt et al., 1982). By contrast, *B. mori* 5.8S RNA shows less conservation than any U-RNA (46% homology with the rat sequence; Fujiwara et al., 1982). As expected, *B. mori* U1 RNA shows excellent primary sequence homology with *Drosophila* U1 RNA (Figure 5).

A clear-cut evolutionary trend for U-RNA primary sequence conservation has not yet emerged in the literature. On the one hand, trends demonstrating a lower sequence conservation in lower organisms can be seen in the following: *Drosophila* U1 shows 72% homology to rat U1 (Mount & Steitz, 1981); *Tetrahymena* and rat U5 show 57% homology (Branlant et al., 1983); *Dictyostelium* and rat U3 show 40% homology (Wise & Weiner, 1980; Takeishi & Kaneda, 1981); and pea plant and rat U5 show 35% homology (Krol et al., 1983). On the other hand, Dinoflagellate U-RNAs show strikingly high homology (80–95%) with vertebrate U-RNAs (Reddy et al., 1983). Interestingly, however, in those domains where primary structure shows poor homology, secondary structure is highly conserved (Krol et al., 1983). The data obtained with *Bombyx* U-RNAs support this generalization (see below).

Secondary Structure of Bombyx U-RNAs. While we have not yet carried out any experiments to probe the secondary structure of *B. mori* U-RNAs, comparisons with published proposed structures for vertebrate U-RNAs shed some light on the evolution of these molecules. Figure 6 illustrates a comparison of *Bombyx* and rat U1 and U5 molecules. For U1 RNA, we used the secondary structure and stem notation described by Mount & Steitz (1981). The rat U1 primary sequence was the corrected one of Branlant et al. (1980). Stems numbered III and IV have been completely sequenced for *Bombyx* U1 and contain six base changes which preserve the hypothesized stem structures. Seven base changes occurred in hypothesized single-stranded domains. Although stems I and II have only been partially sequenced, we observed at least five base changes which preserved the stem II structure. Several base changes and two deletions would destabilize the hypothesized stems. An interesting feature of the stem II and stem III structures in *Bombyx* U1 RNA is that while most base pairs are conserved, several GC base pairs present in the rat sequence are transformed into GU or AU base pairs which are less stable. The same phenomenon occurs in *Drosophila* U1 RNA (Mount & Steitz, 1981).

For molecule U5, we used the secondary structure and stem notation described (Krol et al., 1981; Branlant et al., 1983). We observed a very high base conservation in the long stem region of *Bombyx* U5 and the largest number of base alterations in the loop regions. For this partially sequenced molecule, 71% (15 out of 21) of the base changes occurred in single-stranded domains, which contrasts with the situation observed in U1 where only 30% (8 out of 27) of the base changes occurred in single-stranded regions.

Figure 7 shows structures of *Bombyx* and rat U3 RNAs. For comparison purposes, we used the secondary structure of

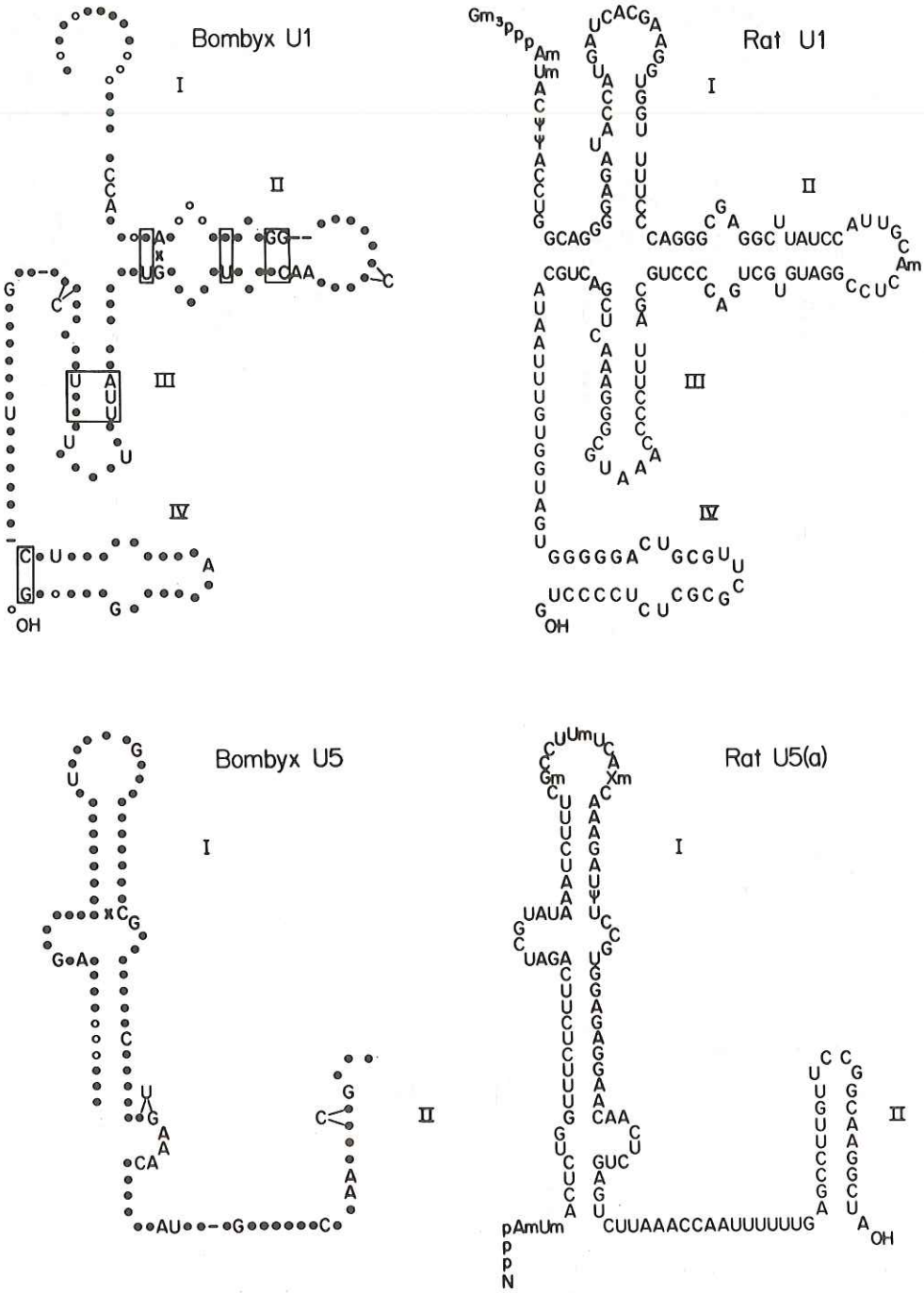

FIGURE 6: Hypothetical secondary structure of *Bombyx* U1 and U5 RNAs. The primary sequences of *Bombyx* U1 and U5 RNAs were determined as described for U3 in Figure 4. Potential secondary structure interactions were identified for each *Bombyx* molecule by using the "RNA-fold" program and were aligned with previously published secondary structures to produce maximum homology. Comparisons of rat and *Bombyx* U1 and U5 are illustrated. Base conservations are indicated by closed circles. Open circles indicate unknown bases. Nonconserved nucleotides are indicated in the *Bombyx* sequence with letters corresponding to the substituted nucleotide. Compensatory base changes which stabilize stem regions are enclosed in boxes. Base changes which destabilize stem regions include an "×" between the bases. For U1 RNA, the secondary structure and stem notation are those described by Steitz and co-workers (Mount & Steitz, 1981). The rat U1 sequence is the modified one of Branlant et al. (1980). For U5 RNA, the secondary structure and stem notation are those of Krol et al. (1981) and Branlant et al. (1983). The rat U5a primary sequence is that of Krol et al. (1981).

rat U3b proposed by Busch and co-workers (Reddy et al., 1979). In *Bombyx*, the most extensive sequencing was done for U3c, the most abundant U3 molecule in Bm-5 cells. Eleven base changes were observed which would conserve the hypothesized stem domains. Nine base changes were observed which would destabilize these regions. Of the base changes, 57% (35 out of 61) occurred in single-stranded domains. For this particular U3, a segment of 11 nucleotides (representing half of the stem III and the adjacent single-stranded region) has been deleted. This deletion is not observed for U3a or U3b. Molecule U3d has not been sequenced through stem III yet. Notice that the oligo(G) stretch corresponding to nucleotides 31–35 in the double-stranded region of stem II for U3c is strikingly underrepresented (not cleaved efficiently) in the enzymatic guanosine reaction of the sequencing gel (Figure 4, lanes 4 and 5). This observation suggests that the stem II structure may indeed form in *B. mori* U3 RNA. However, the extensive primary sequence changes in U3c and U3d relative to rat U3 cast doubt on whether other proposed features of the vertebrate molecule are conserved in its insect

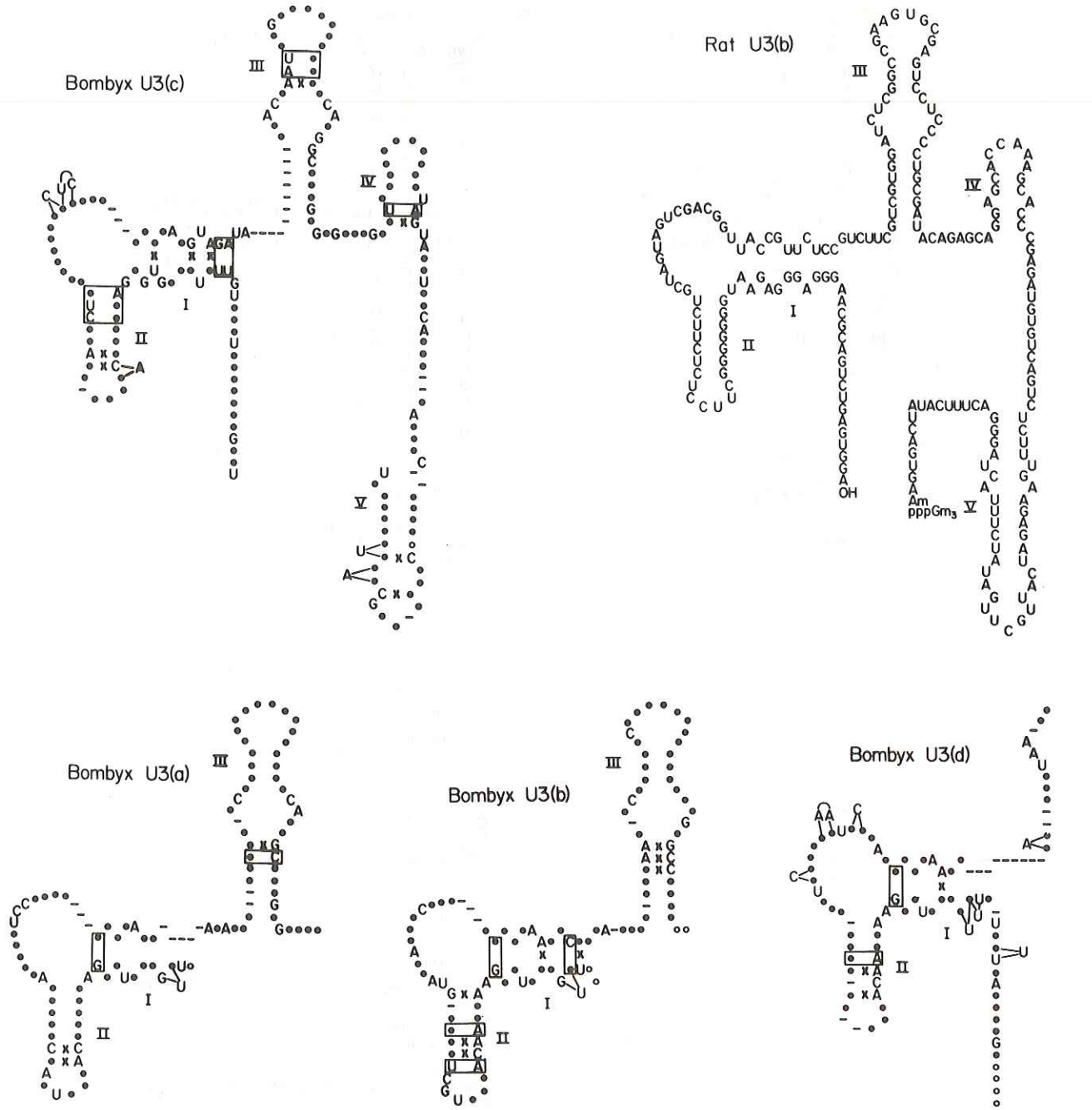

FIGURE 7: Hypothetical secondary structure of *Bombyx* U3 RNAs. All details are as given in Figure 6. Comparisons of rat U3b and *Bombyx* U3 (a, b, c, and d) are shown. The primary and secondary rat U3b structures are those of Reddy et al. (1979).

counterparts.

In general, our data suggest that with the possible exception of U3 RNA, secondary structure may be highly conserved between vertebrate and *Bombyx* U-RNAs. High degrees of secondary structure conservation for U-RNAs may indicate a high level of functional constraint for these molecules and supports the hypothesis that they play fundamental cellular roles for which such structures are essential. The recent finding (Guerrier-Takada et al., 1983) that the catalytic subunit of the RNA processing enzyme RNase P is an RNA molecule is relevant in this respect. Since some of the U-RNAs have been proposed to be components of the cell's RNA processing machinery [see Busch et al. (1982) for a review], the high degree of structural conservation in evolution could perhaps reflect the constraints of an underlying catalytic (ribozymic) function.

ACKNOWLEDGMENTS

We thank T. Genetta for technical assistance and R. Binder for help in the RNA sequencing experiments. We are indebted to Drs. James Vaughn and Dennis Knudsen of the Insect Pathology Laboratory of the U.S. Department of Agriculture, who provided us with inocula of the Bm-5 cell line. HeLa cells were kindly provided by Dr. Joseph Nevins of The Rockefeller University. The manuscript was expertly typed by I. Guzman.

REFERENCES

Adams, D. S., Luhrmann, R., & Lizardi, P. M. (1984) *Gene Anal. Tech. 1*, 109–115.

Bester, A. J., Kennedy, D. S., & Heywood, S. M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1523–1527.

Bonner, W. M., & Laskey, R. A. (1974) *Eur. J. Biochem. 46*, 83–88.

Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J., & Jeanteur, P. (1980) *Nucleic Acids Res. 8*, 4143–4154.

Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Dias, L., & Pousada, C. (1983) *Nucleic Acids Res. 11*, 8359–8367.

Bringmann, P., Reuter, R., Rinke, J., Appel, B., Bald, R., & Luhrmann, R. (1983) *J. Biol. Chem. 258*, 2745–2747.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., & Luhrmann, R. (1984) *EMBO J. 3*, 1357–1363.

Busch, H., Reddy, R., Rothblum, L., & Choi, Y. C. (1982) *Annu. Rev. Biochem. 51*, 617–654.

Calvet, J. P., & Pederson, T. (1981) *Cell (Cambridge, Mass.) 26*, 363–370.

Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527–2538.

England, R. E., & Uhlenbeck, O. C. (1978) *Nature (London) 275*, 560–561.

Forbes, D. J., Kornberg, T. B., & Kirschner, M. W. (1983) *J. Cell Biol. 97*, 62–72.

Fujiwara, H., Kawata, Y., & Ishikawa, H. (1982) *Nucleic Acids Res. 10*, 2415–2418.

Galli, G., Hofstetter, H., Stunnenberg, H. G., & Birnstiel, M. L. (1983) *Cell (Cambridge, Mass.) 34*, 823–828.

Gardiner, G. R., & Stockdale, H. (1975) *J. Invertebr. Pathol. 25*, 363–370.

Garvey, J. S., Cremer, N. R., & Sussdorf, D. H. (1977) *Methods in Immunology*, 3rd ed., W. A. Benjamin, Reading, MA.

Gilbert, W., & Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3581–3584.

Goldstein, L. (1976) *Nature (London) 261*, 519–521.

Grace, T. D. C. (1967) *Nature (London) 216*, 613.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983) *Cell (Cambridge, Mass.) 35*, 849–857.

Harada, F., & Kato, N. (1980) *Nucleic Acids Res. 8*, 1273–1285.

Hashimoto, C., & Steitz, J. A. (1984) *Nucleic Acids Res. 12*, 3283–3293.

Hellung-Larsen, P., & Frederiksen, S. (1977) *Comp. Biochem. Physiol. B 58B*, 273–281.

Ireland, L., Szyszko, J., & Krause, M. (1982) *Mol. Biol. Rep. 8*, 97–101.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. B., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1398–1402.

Kanehisa, T., Oki, Y., & Ikuta, K. (1974) *Arch. Biochem. Biophys. 165*, 146–152.

Kessler, S. W. (1975) *J. Immunol. 115*, 1617–1624.

Krol, A., Gallinaro, H., Lazar, E., Jacob, M., & Branlant, C. (1981) *Nucleic Acids Res. 9*, 769–787.

Krol, A., Ebel, J.-P., Rinke, J., & Luhrmann, R. (1983) *Nucleic Acids Res. 11*, 8583–8594.

Lehrach, H., Diamond, D., Wozney, J. M., & Boedtker, H. (1977) *Biochemistry 16*, 4743–4750.

Lerner, E. A., Lerner, M. R., Janeway, C. A., & Steitz, J. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2737–2741.

Lerner, M. R., & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5495–5499.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., & Steitz, J. A. (1980) *Nature (London) 283*, 220–224.

Lizardi, P. M., & Engelberg, A. (1979) *Anal. Biochem. 98*, 116–122.

Luhrmann, R., Appel, B., Bringmann, P., Rinke, J., Reuter, R., & Rothe, S. (1982) *Nucleic Acids Res. 10*, 7103–7113.

Mount, S. M., & Steitz, J. A. (1981) *Nucleic Acids Res. 9*, 6351–6368.

Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., & Steitz, J. A. (1983) *Cell (Cambridge, Mass.) 33*, 509–518.

Murray, V., & Holliday, R. (1979) *FEBS Lett. 106*, 5–7.

Padgett, R. A., Mount, S. M., Steitz, J. A., & Sharp, P. A. (1983) *Cell (Cambridge, Mass.) 35*, 101–107.

Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1760–1764.

Pederson, T., & Bhorjee, J. S. (1979) *J. Mol. Biol. 128*, 451–480.

Penman, S. (1966) *J. Mol. Biol. 17*, 117–130.

Reddy, R., Henning, D., & Busch, H. (1979) *J. Biol. Chem. 254*, 11097–11105.

Reddy, R., Spector, D., Henning, D., Liu, M.-H., & Busch, H. (1983) *J. Biol. Chem. 258*, 13965–13969.

Ringuette, M., Liu, W.-C., Jay, E., Yu, K. K., & Krause, M. O. (1980) *Gene 8*, 211–224.

Rinke, J., Appel, B., Blocker, H., Frank, R., & Luhrmann, R. (1984) *Nucleic Acids Res. 12*, 4111–4126.

Rogers, J., & Wall, R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1877–1879.

Smith, J. H., & Eliceiri, G. L. (1983) *J. Biol. Chem. 258*, 4636–4640.

Sohn, U., Szyszko, J., Coombs, D., & Krause, M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 7090–7094.

Suzuki, Y., & Brown, D. D. (1972) *J. Mol. Biol. 63*, 409–429.

Takeishi, K., & Kaneda, S. (1981) *J. Biochem. (Tokyo) 90*, 299–308.

Thimmappaya, B., Weinberger, C., Schneider, R. J., & Shenk, T. (1982) *Cell (Cambridge, Mass.) 31*, 543–551.

Tinoco, I., Borer, P., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature (London), New Biol. 246*, 40–41.

Troutt, A., Savin, T. J., Curtiss, W. C., Celentano, J., & Vournakis, J. N. (1982) *Nucleic Acids Res. 10*, 653–664.

Vournakis, J. N., Celantano, J., Finn, M., Lockard, R. E., Mitra, T., Pavlakis, G., Troutt, A., Van den Berg, M., & Wurst, R. M. (1981) in *Gene Amplification and Analysis* (Chirikjian & Papas, Eds.) Vol. 2, pp 267–298, Elsevier/North-Holland, New York.

Walter, P., & Blobel, G. (1982) *Nature (London) 299*, 691–698.

Wieben, E. D., Madore, S. J., & Pederson, T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1217–1220.

Wise, J. A., & Weiner, A. M. (1980) *Cell (Cambridge, Mass.) 22*, 109–118.

Wise, J. A., & Weiner, A. M. (1981) *J. Biol. Chem. 256*, 956–963.

Wise, J. A., Tollervey, D., Maloney, D., Swerdlow, H., Dunn, E. J., & Guthrie, C. (1983) *Cell (Cambridge, Mass.) 35*, 743–751.

Wooley, J. C., Cone, R. D., Tartof, D., & Chung, S.-Y. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6762–6766.

Yang, V. W., Lerner, M. R., Steitz, J. A., & Flint, S. J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1371–1375.

Zeller, R., Nyffenegger, T., & DeRobertis, E. M. (1983) *Cell (Cambridge, Mass.) 32*, 425–434.

Zuker, M., & Stiegler, P. (1981) *Nucleic Acids* Res. *9*, 133–148.