Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants

Andreas Woppmann, Thilo Patschinsky¹, Peter Bringmann², Franz Godt³ and Reinhard Lührmann*

Institut für Molekularbiologie und Tumorforschung der Philipps-Universität Marburg, Emil-Mannkopf-Straße 2, D-3550 Marburg, ¹Heinrich-Pette-Institut, Martinistraße 52, D-2000 Hamburg 20, ²Schering AG, Müllerstraße 170–178, D-1000 Berlin 65, ³Technische Fachhochschule Berlin, Seestraße 10, D-1000 Berlin 65, FRG

Received May 11, 1990; Revised and Accepted June 13, 1990

ABSTRACT

The proteins of the major human snRNPs U1, U2, U4/U6 and U5 were characterised by two-dimensional electrophoresis, with isoelectric focussing in the first dimension and SDS-polyacrylamide gel electrophoresis in the second. With the exception of protein F, which exhibits an acidic pI value (pI = 3.3), the snRNP proteins are basic. Post-translational modification was found among the proteins associated specifically with the U1 and U2 particles. The most complex modification pattern was observed for the U1-specific 70K protein. This was found in at least 13 isoelectric variants, with pI values ranging from 6.7 to 8.7; these variants differed also in molecular weight. All of the 70K variants are phosphorylated in the cell. Thin-layer analysis of their tryptic phosphopeptides revealed that the 70K variants have four major phosphopeptides in common, in addition to which at least four additional serine residues are phosphorylated to different extents. The comparative phosphopeptide analysis shows that differential phosphorylation alone is not sufficient to explain the occurrence of the many isoelectric variants of 70K, so that the final charge of the 70K variants is determined both by phosphorylation and by other, as yet unidentified posttranslational modifications. By two-dimensional separation of snRNP proteins obtained from mouse Ehrlich ascites tumour cells, it was shown that the pattern of pI values of the mouse proteins was almost identical with the corresponding pattern for human proteins. Even the complex modification patterns of the 70K protein are identical in mouse and man, indicating that the presence in the cell of so many variants of this protein may have functional importance. The major difference between murine and human snRNP proteins is the absence of protein B' from mouse snRNPs. This suggests that the

homologous protein B may be able to carry out the task of protein B'.

INTRODUCTION

Eucaryotic cells contain a group of small nuclear RNAs, the snRNAs U1, U2, U4, U5 and U6. In the nucleoplasm, these are organised as four discrete RNP particles, the snRNPs U1, U2, U4/U6 and U5 (1), and there is experimental evidence that all these RNPs participate in the splicing of pre-mRNA (2, 3). One of their functions appears to be the recognition of certain signal structures in the pre-RNA molecule, such as the 5' and 3' splicing sites, or the branching point (3, 5). The importance of the protein components of the snRNPs for modulating the function of these particles has been underlined by the finding that snRNP proteins are required for the efficient formation *in vitro* of the complex between U1 RNA and 5' splicing junctions (6, 7).

The proteins of the snRNPs may be divided into two classes: the common proteins B', B, D, D', E, F and G, which are common to all snRNPs, and other proteins that are specific for particular snRNPs. Thus, U1 snRNPs possess in addition to the common proteins the specific proteins 70K, A and C, while U2 contains the specific proteins A' and B'' (8). U5 snRNP is the snRNP particle with the greatest proportion of polypeptides; it contains at least seven U5-specific proteins, characterised by molecular masses of 40, 52, 100, 102, 116 and 200 kDa, the latter usually appearing as a double band in electrophoresis (9).

The snRNP proteins are also interesting from a clinical and an immunological standpoint. Patients with systemic lupus erythematosus or related connective tissue diseases often develop autoantibodies that react with particular snRNP proteins (10, 11). Anti-(U1)RNP autoantibodies react with the U1 polypeptides 70K, A and C, so that they precipitate only U1 snRNPs. In contrast, anti-Sm autoantibodies precipitate all the nucleoplasmic snRNPs, as the major immunoreactive Sm proteins are the

* To whom correspondence should be addressed

proteins B, B' and D, common to all snRNPs (12). Autoantibodies against the U2 RNP-specific polypeptides A' and B'' as well as against the common proteins E, F and G also occur in the sera of some patients, but they occur less frequently (13, 14).

Molecular cloning and sequence analysis of the cDNAs for several human snRNP proteins (70K, A, B'', A', C, B', B, D and E) have yielded information about the primary structure of these proteins (15). Little is known, however, about the post-translational modification of the snRNP proteins. Yet accurate characterisation of the post-translational state of the proteins is important for two reasons. (i) Post-translational modifications of a particular snRNP protein could be responsible for the regulation of the activity of the snRNP with which the protein is associated, or for changing the specificity of interaction of the snRNP with a pre-mRNA sequence. (ii) The snRNP particles and their proteins may well act as immunogens in the anti-snRNP autoantibody response (16), and it is possible that aberrant post-translational modification of RNP or Sm-antigenic proteins may render these proteins foreign to the individual's own immune system and thus initiate an autoimmune response.

As one step toward this goal, we have characterised the snRNP proteins from man and mouse by two-dimensional electrophoresis, using isoelectric focussing in the first dimension and SDS-polyacrylamide gel electrophoresis in the second. With the sole exception of protein F, which has an acidic pI value (pI = 3.3), all the snRNP proteins are basic. Post-translationally modified proteins were found among the U1- and U2-specific polypeptides. The most complex modification pattern was found for the U1-specific 70K protein, which occurs in at least twelve variants differing both in electric charge (pI = 6.7–8.7) and in molecular weight. We have carried out 2-D thin-layer analysis of the phosphopeptides obtained by tryptic digestion of these variants. This showed that although heterogeneity with respect to phosphorylation does contribute to the number of variants, other modifications as yet unidentified must also contribute to this.

MATERIALS AND METHODS

Preparation of snRNPs and snRNP proteins

The cultivation of HeLa S3 cells or Ehrlich ascites tumour cells in suspension culture, their labelling with [³²P]-orthophosphate and the preparation of nucleoplasmic extracts (termed NX-50, because extracts were prepared at 50 mM MgCl₂ concentration) were carried out in the manner described by Bringmann *et al.* (17). snRNPs U1–U6 were isolated from nuclear extracts by immunoaffinity chromatography with the monoclonal anti-m₃G antibody H20 (18) by a method based upon that of Bringmann *et al.* (17). In all experiments, the snRNPs were dialysed before use against a buffer containing 5 mM (Hepes) at pH 7.9, 15% glycerol, 25 mM KCl, 1 mM MgCl₂, 0.05 mM ethylene diamine tetra-acetic acid (EDTA), 0.5 mM dithioerythritol (DTE) and 0.5 mM phenyl methyl sulphonyl fluoride (PMSF).

For each sample to be analysed by 2-D electrophoresis, 50 µg snRNP proteins (0.1–0.2 mg/ml) were extracted with PCA (phenol/chloroform/*iso*-amyl alcohol in the proportions 50:49:1 v/v) on an Eppendorf shaker for 1 h at 37°C. After centrifugation, the organic phase and the interface were carefully removed, and the proteins were precipitated from this by the addition of a 5-fold excess of acetone followed by chilling to –20 or –80°C. The pellets were washed in 80% ethanol (–20°C) and dried *in vacuo*.

Two-dimensional protein analysis

1st. dimension: Electrophoresis was performed by a modification of standard methods (19–21). The cylindrical gels (0.15×12 cm) consisted of an 11 cm separation gel (84% w/v acrylamide, 0.3% w/v bisacrylamide) and a 1 cm cap gel (10% w/v acrylamide, 0.13% w/v Bis), both containing 9 M urea and 1% v/v carrier ampholytes. For isoelectric focussing (IEF), Servalyt 3-10 and Servalyt 9-11 were mixed in the ratio 4:1. For non-equilibrium pH-gradient electrophoresis (NEPHGE), Servalyt 3-10 was used unmodified.

For electrophoresis, the cylindrical gels in glass capillaries were mounted vertically (apparatus from Shandon, UK) with the cap gel at the bottom. A 1 mm layer of Sephadex G200 (suspended in 9 M urea, 5% v/v 2-mercaptoethanol and 1% v/v ampholyte; the Sephadex was first hydrated in 25% sucrose) was placed on the top of the gel to prevent clogging by undissolved material. After dissolution in 10 µl sample buffer (9 M urea, 5% v/v 2-mercaptoethanol, 2% v/v ampholyte, 16% w/v glycerol) the protein samples were applied to the Sephadex suspension and covered with overlay solution (5 M urea, 5% v/v glycerol, 1% v/v ampholyte). The buffer in the upper chamber was 0.75 M H₃PO₄ (adjusted to pH 1.2 with NaOH) and the buffer in the lower chamber 0.75 M ethylene diamine (adjusted to pH 12.7 with HCl). The buffer in the upper chamber contained in addition 3 M urea and that in the lower chamber 9 M urea and 5% w/v glycerol.

IEF was carried out at room temperature (normally *ca.* 24°C) with the anode at the top and the cathode at the bottom, with the voltage regulated as follows: 1 h at 100 V, 45 min at 400 V, 3 h at 600 V, 30 min at 800 V, 30 min at 1000 V, 30 min at 1500 V, 5 min at 2000 V, giving 4000 volt-hours in all. For NEPHGE the direction of the applied voltage was reversed and the regulation was: 1 h at 100 V, 1 h 40 min at 300 V, 6 min at 600 V, giving 700 volt-hours. The pH gradients were calibrated by using pH-marker proteins ('Broad pI' and 'High pI', from Pharmacia) in parallel cylindrical gels and staining these with Coomassie Blue R250.

2nd. dimension: This was resolved by standard methods of discontinuous SDS-polyacrylamide gel electrophoresis (22) employing 20×20×0.13 cm gels with 7% or 15% polyacrylamide as required.

The cylindrical gels were removed from their glass tubes and equilibrated (15 minutes) with 4×stacking-gel buffer (250 mM Tris.HCl pH 6.8, 0.4% w/v SDS). They were then placed on the stacking gel (2 cm above the junction with the separation gel) and polymerised into it by the addition of a further 1 cm stacking gel. The electrophoresis was followed with bromphenol blue as tracking dye, applied in Laemmli sample buffer above the polymerised gel. Electrophoresis was carried out overnight at 50 V, followed by 30 min at 150 V and finally 300 V until the bromphenol blue ran out at the bottom of the gel.

Immunoblotting

Total snRNP protein was separated by 2-D gel electrophoresis with IEF in the first dimension and electrophoresis in a 15% SDS-polyacrylamide gel in the second. The proteins were transferred electrophoretically onto nitrocellulose (23). Probing of snRNP proteins with anti-(U1)RNP autoimmune sera or with affinity-purified anti-70K autoantibodies was carried out essentially as described elsewhere (16).

Affinity purification of anti-70K antibodies

E. coli cells transformed with plasmid p70.1, which codes for a *cro*- β -galactosidase-70K fusion protein (24) were grown in Luria-Bertain medium overnight at 30°C. The temperature of the cultures was then raised to 42°C and incubation was continued for another 2 h in order for expression of the fusion protein to take place. The cells were pelleted by centrifugation, lysed with lysozyme and then treated with Triton X-100 in order to bring about enrichment of the hybrid proteins. The proteins insoluble in Triton X-100 were subjected to preparative electrophoresis on 8% SDS-polyacrylamide gels and then transferred electrophoretically onto nitrocellulose. The blots were allowed

to react with an anti-(U1)RNP autoimmune serum containing anti-70K antibodies. The region of the blot containing the *cro*- β -galactosidase-70K fusion protein was excised and the bound antibodies were eluted (25).

Total tryptic proteolysis of 32 P labelled 70K protein

snRNPs U1 to U6 were isolated by anti-m₃G immunoaffinity purification from nuclear extracts NX-50 prepared from HeLa cells which had been grown in medium containing 32 P orthophosphate. Following fractionation of snRNP proteins in a preparative 15% SDS polyacrylamide gel the 32 P-labelled 70K protein was eluted from the gel (26), lyophilised and freed from

Fig. 1. pH gradient of the IEF dimension. For the calibration of the pH gradient in the isoelectric-focussing dimension, the pI-marker protein mixtures 'high pI' (hpl) and 'broad pI' (bpl) (Pharmacia) were analysed in parallel with the snRNP proteins. (A): After IEF the positions of the pI markers were determined by Coomassie Blue staining. The pI values of the marker proteins stated by the supplier were: 1, 3.50; 2, 4.55; 3, 5.20; 4, 5.85; 5, 6.55; 6, 6.85; 7, 7.35; 8, 8.15; 9, 8.45; 10, 8.65; 11, 9.30; 12, 10.25. (B): Plot of the stated pI values of the marker proteins against their relative distance from the point of application. (C): Projection onto the abscissa of the plot in (B). This was used to calibrate the pH gradient in IEF runs.

residual SDS and salt by dissolving them in 400 μ l water, centrifuging for 15 min and precipitating them from the supernatant with trichloroacetic acid (TCA; to 20% w/v, overnight at 4°C). The proteins were washed at -20°C with absolute ethanol, then with ethanol/diethyl ether (1:1, freshly mixed) and were dried under nitrogen.

Performic acid was prepared by mixing 9 parts v/v formic acid with 1 part concentrated hydrogen peroxide. The proteins, purified as above, were incubated with 50 μ l fresh performic acid at room temperature for 2 h and were then diluted with 1 ml water, quench-frozen and freeze-dried. The oxidised proteins were resuspended in 50 μ l at 50 mM ammonium bicarbonate.

In the first step of proteolysis, the resuspended proteins were incubated with 30 μ g TPCK-trypsin (1 mg/ml in 50 mM ammonium bicarbonate) for 4 h at 37°C. A further 20- μ g portion of TPCK-trypsin was added and the incubation was continued

for another 2 h. The samples were diluted with 1 ml water and freeze-dried. They were then dissolved in 500 μ l water and freeze-dried; this step was carried out four times in order to remove all the ammonium bicarbonate. Before the final lyophilisation, the samples were centrifuged in order to remove any insoluble particles and again freeze-dried. The tryptic peptides were stored at -80°C.

Phosphopeptide analysis

Phosphoproteins prepared by polyacrylamide gel electrophoresis were subjected to total proteolysis as described in the previous paragraph. The peptides were taken up in 10–20 μ l water and applied to cellulose-coated glass plates (20×20 cm) for thin-layer electrophoresis (26). The point of application was 10.5 cm from the left-hand edge and 2.5 cm from the bottom edge. Electrophoresis (anode, left-hand edge; cathode, right-hand edge)

Fig. 2. 2-D analysis of HeLa snRNP proteins. (A): 50 μ g purified U1-U6 snRNP proteins from HeLa cells were separated by isoelectric focussing in a gradient from pH 3 to pH 11 for 4000 V-h. The pH gradient was charted by use of pI markers (Figure 1). The second dimension was run in 15% polyacrylamide with SDS. For details, see Materials and Methods. Carbamylated glyceraldehyde phosphate dehydrogenase (native pI 8.5) was added to the sample as an internal standard. To ensure correct identification of the Coomassie-stained proteins, 20 μ g snRNP proteins and molecular-weight standards were run in separate lanes of the SDS-polyacrylamide gel electrophoresis. The molecular weights of the standards were: 92.5, 62.5, 45, 31, 21.5, and 14.4 kDa (from top to bottom). (B): 30 μ g snRNP proteins were applied at the cathode end in the first dimension and separated by migration toward the anode. After non-equilibrium pH gradient electrophoresis (NEPHGE) for 700 V-h, the analysis was continued with SDS-polyacrylamide gel electrophoresis in the second dimension. Only the acidic protein F is seen.

Table 1

snRNP proteins

Protein	Apparent MW (kDa)	Isoelectric variant no.	pI (measured by IEF)
70K	70	1	6.77
		2	6.88
		3	7.08
		4	7.12
		5	7.33
		6	7.44
		7	7.77
		8	7.82
		9	8.05
		10	8.18
		11	8.34
		12	8.52
		13	8.62
A	34		10.0
A'	33	1	8.3
		2	8.6
B'	29		10.7
B''	28.5		10.2
B	28		10.6
C	22		10.4
D	16		10.5
D'	15.5		8.0
E	12		10.3
F	11		≤3.3
G	9		8.8

Table 1. pI values of the HeLa snRNP proteins and their isoelectric variants. By determination of the distances of the snRNP proteins from the anodic end of the IEF dimension, the isoelectric points of the snRNP proteins could be read off from the respective pH diagrams (compare Figure 1B). The snRNP proteins and their isoelectric variants are shown along with their molecular weights and isoelectric points.

was carried out at 1300 V for 30 min in a buffer consisting of 1% w/v ammonium carbonate (pH measured, 8.9). The second dimension of resolution was attained by chromatography (5–5.5 h) in phosphopeptide chromatography buffer, which was *n*-butanol/pyridine/acetic acid/water in the ratio 75:50 : 15:60 v/v (pH measured, 5.2). The phosphopeptides were detected by autoradiography.

RESULTS

Two-dimensional separation of HeLa snRNP proteins

The mixture of snRNPs U1, U2, U4/U6 and U5 was prepared by extraction of nuclei with a buffer containing 0.5 M NaCl and 50 mM MgCl₂ (17, 27). This extract, termed NX-50, was used as the source of snRNP proteins. These were purified by anti-m₃G immuno affinity chromatography. As described in previous publications (17, 28), snRNPs prepared in this way contain the common proteins B', B, D, D', E, F and G and also the U1- and U2-specific proteins 70K, A, C, B'' and A', but they lack the U5-specific proteins (40, 52, 100, 102, 116 and 200 kDa). The latter can be isolated as constituents of 20S U5 snRNP particles, if the snRNPs U1 to U6 are purified from splicing extracts, *i.e.*, nuclear extracts prepared using a low-salt buffer (9). However, we have as yet failed to obtain reproducible two-

dimensional separations of the U5-specific proteins. Therefore, in this study, snRNPs purified from NX-50 extracts were used as starting material, so that any obscuring of cleanly-running snRNP proteins by streaks from U5-specific proteins could be avoided. This also appeared to be warranted in so far as the electrophoretic behaviour of the common proteins and of the U1- and U2-specific proteins appeared to be completely independent of whether the snRNPs had been isolated from NX-50 or from splicing extracts (data not shown).

The snRNP proteins were separated in the first dimension either by isoelectric focussing (IEF) or by non-equilibrium pH-gradient electrophoresis (NEPHGE). Best reproducibility in the first (IEF) dimension was obtained when the ampholyte was a 4:1 mixture of Servalyt 3–10 and Servalyt 9–11 and with electrofocussing carried out for 4000 V-h. Under these conditions, a pH gradient ranging from pH 3 to pH 11 with an almost constant slope between pH 6 and 9 was produced, as shown by pI marker proteins (Fig. 1). In the second dimension, the proteins were fractionated in a 15% SDS-polyacrylamide gel, except in some experiments where optimal resolution of the 70K proteins was required, in which case 7% polyacrylamide was used.

Figure 2 depicts a typical analysis, using IEF in the first dimension, of the snRNP proteins contained in the mixture of snRNPs U1, U2, U4/U6 and U5 from HeLa extracts. The pI values of the snRNP proteins and their isoelectric variants are summarised in Table 1; here and in the rest of this paper, the term 'pI value' refers to the value as found by calibration of the pH gradients as described in Materials and Methods. Protein F is the only highly acidic protein, with a pI value of about 3.3. If analysed under NEPHGE conditions, *i.e.*, with migration from cathode to anode, protein F migrates in front of all the other polypeptides and has already passed two thirds of the pH gradient gel after 700 V-h (Fig. 2B).

With the further exception of several variants of the 70K protein (see below), all the other snRNP proteins are of basic character. The common proteins B', B and D are the most basic ones, with pI values between 10.5 and 10.7 (Table 1).

Some proteins show unusual electrophoretic properties, particularly during the isoelectric-focussing procedure. At a molecular weight of about 16 kDa, two protein spots are resolved, the major one with a pI value of 10.5 and the minor one with pI 8.0 (Fig. 2A). Its reactivity with anti-D antibodies in immunoblots has led to the conclusion that the major spot contains protein D. In the same experiments, the minor D-sized protein species reacted only weakly or not at all, depending on the autoimmune serum used (not shown). It is therefore unlikely that this protein represents a post-translationally modified D variant, so we tentatively term this protein D'. In an earlier paper, we have described the appearance of an additional D-sized protein in one-dimensional SDS-polyacrylamide gel electrophoresis, which we also termed D' (28). However, we do not yet have concrete evidence that these are one and the same protein, and this question must remain an open one for the time being.

The U1-specific C protein sometimes separates in two faint spots, slightly differing in molecular weight; this behaviour has been observed before on one-dimensional SDS-polyacrylamide gels (29).

Polypeptides A and B'' sometimes produce a broad smear during the isoelectric-focussing step in addition to the discrete spots at pI values 10 and 10.2, respectively. It is not yet clear whether this propensity to aggregate is due to post-translational modification of the two proteins.

Fig. 3. Immunodetection of protein 70K. Four samples of HeLa snRNP proteins (50 μ g each) were analysed in parallel in the manner of Figure 2A. One of the 2-D gels was stained with Coomassie Blue (A). The 70K regions of the other three gels were excised as indicated in (A) and the proteins were transferred electrophoretically onto nitrocellulose strips. The efficiency of the transfer was checked by the Poinceau method (B). Finally, the nitrocellulose strips were subjected to immuno-affinity staining, as follows: NHS, normal human serum; α -RNP, anti-(U1)RNP serum; α -70K, antibodies specific against a fusion protein of 70K and β -galactosidase (see text).

The U2-specific protein A' separates reproducibly into two isoelectric variants, a major one with a pI value of 8.6 and a minor, more acidic one with a pI value of 8.3.

However, the most interesting electrophoretic behaviour is observed for the U1-specific 70K protein. At least 11, sometimes as many as 13 isoelectric variants can be distinguished, separating with pI values from 6.7 to 8.7. (The minor basic variants 12 and 13 can only be seen if greater amounts of snRNP proteins are used, as for example in Fig. 3). It should be noted that the pI values of the 70K variants are quite irregularly distributed. Thus, some 70K variants, *e.g.*, 3/4, 5/6 and 7/8, focus in pairs

with only slight differences in pI, while others, *e.g.*, variants 8 and 9, show no such regularity, with gaps of 0.2 pI units or more (see Fig. 2A and Table 1).

Reactivity of all 70K variants with anti-U1-70K antibodies

Experimental evidence has been adduced for the existence of an snRNP protein with a molecular weight (as observed on SDS-polyacrylamide gels) of about 70 kDa, which, however, could be distinguished from the U1-specific 70K protein by its ability to react with anti-Sm antibodies (30). We therefore investigated whether all of the 13 isoelectric variants with molecular weight

70 kDa that we could resolve in the 2-D gel electrophoresis were genuine variants of U1-specific 70K. This was assessed by two-dimensional immunoblots with anti-U1-70K autoantibodies.

Figure 3A shows the Coomassie-stained proteins of the U1-U6 snRNP preparation used for the immunoblot experiments, after two-dimensional separation. This preparation contained at least one additional protein with showing a molecular weight of about 68 kDa, lying alongside the 70K variants and focussing at pI *ca.* 9.8.

For the following immunoblot experiments (Fig. 3B), we depict only the sections of the 2-D gel containing the 70K variants, as indicated in Figure 3A. The transfer efficiency on the

nitrocellulose membrane of all 70K variants is confirmed by staining with Ponceau S stain (Figure 3B).

In Figure 3C it is demonstrated that all proteins found in the pI range 6.7–8.7 actually reacted with an anti-(U1)RNP serum (middle strip) that possessed a high titre against the U1-specific 70K protein and precipitated exclusively U1 snRNPs (data not shown). However, the 68-kDa protein with a pI value of 9.8 did not react with this anti-(U1)RNP serum. A normal human serum that did not contain anti-snRNP autoantibodies showed no reactivity with the 70K variants (Figure 3C, top strip). In order to exclude the possibility that the staining of some of the 70K proteins might be due to a slight contamination of the anti-

Fig. 4. 2-D analysis of the U1 snRNP protein 70EK. 50 μ g U1-U6 snRNP proteins from HeLa cells were analysed as shown in Figure 2A, with the difference that the SDS-polyacrylamide gel electrophoresis was performed in 7% polyacrylamide gels, since this allows better separation of the 70K variants on the basis of slight differences in molecular weight.

Fig. 5. 2-D analysis of snRNP proteins from Ehrlich ascites tumour cells. 50 μ g of purified U1-U6 snRNP proteins from Ehrlich ascites tumour cells were separated by IEF in the first dimension and by 15% SDS-polyacrylamide gel electrophoresis in the second, in a manner analogous to that employed in Figure 2A.

Fig. 6. 2-D analysis of the U1 snRNP-phosphoprotein 70K. PCA-extracted proteins (see Materials and Methods) from 50 μ g of unlabelled U1-U6 snRNPs (A) and from 2×10^6 cpm 32 P-labelled U1-U6 snRNPs (B) were separated by IEF in a gradient from pH 3 to pH 11 in the first dimension followed by electrophoresis in a 15% SDS-polyacrylamide gel in the second. The molecular weight range shown is approximately 45–100 kDa. The positions of the isoelectric variants 1 to 11 are indicated.

(U1)RNP serum with populations of anti-Sm autoantibodies, we isolated the anti-70K autoantibodies from the serum by immunoaffinity purification with recombinant 70K protein. Figure 3C (lowest strip) proves that the highly specific anti-70K autoantibodies also react with all 70K variants.

Size differences among some 70K variants

Close inspection of the 2-D gels of the snRNP proteins reveals slight differences in the migration behaviour of several 70K variants (for example, variants 1, 2 and 8) when these are separated in a 15% SDS-polyacrylamide gel in the second dimension. In order to investigate this further, we repeated the two-dimensional separation, this time using a 7% SDS-polyacrylamide gel for the fractionation of the proteins in the second dimension. Under these conditions, differences in migration behaviour of 70-kDa polypeptides become more pronounced. It was found that variants 1 and 2 of the HeLa 70K protein migrated substantially more slowly than the other variants did (Figure 4). Differences in migration behaviour can also be observed between variants lying close together, such as the 'pairs' 5/6 and 7/8. These results show clearly that some of the 70K variants differ not only in pI value but also in their apparent molecular weight.

Two-dimensional separation of mouse snRNP proteins

Figure 5 shows a two-dimensional separation of the protein from the snRNPs U1, U2, U4/U6 and U5, which had been isolated from mouse Ehrlich ascites tumour cells by the procedure described for HeLa cells above. The correspondence between the 2-D patterns of mouse and human snRNP proteins is striking. This applies not only to the common proteins B', B, D, D', E, F and G, but also to the U1- and U2-specific proteins 70K, A, C, A' and B''. It is especially noteworthy that both the number and the pI values of the isoelectric variants of the proteins A' and 70K as observed under these conditions are identical in mouse and man. A clear difference can be noticed with regard to protein B': mouse snRNPs show a single Coomassie-stained protein spot

with a molecular weight of 28 kDa and a pI value of 10.5, which correspond to the physico-chemical characteristics of human protein B. It is therefore probable that protein B' is not present in mouse snRNPs.

Phosphorylation of all 70K variants *in vivo*

It has been shown previously that the U1-specific 70K protein is the only snRNP protein to be phosphorylated *in vivo*, and that this occurs exclusively at serine residues (31). We were therefore interested in investigating the state of phosphorylation of the 70K variants, and also in finding out whether differential phosphorylation contributes to the occurrence of the many 70K variants.

Figure 6 shows the autoradiogram of the 70K region of the 2-D analysis of a protein mixture of snRNPs that were isolated from HeLa cells after labelling *in vivo* with 32 P-phosphate. Because of the long exposure, the resolution of the 70K variants in the autoradiogram is not as exact as in the corresponding Coomassie blue-stained SDS-polyacrylamide gel; however, it can clearly be seen that all of the variants are phosphorylated. In addition, comparison of the two parts of Figure 6 show that the intensity of the blackening of the individual protein spots in the autoradiogram, that is, the extent of the phosphorylation, is correlated with the mass of the protein of the corresponding 70K variants in the SDS-polyacrylamide gel as measured by the intensity of the Coomassie staining.

Phosphopeptide analysis of 70K variants

More precise information about the differences in the position of phosphorylation of the 70K variants was expected from the analysis of the number of phosphorylation sites in the various variants. For this purpose, the 70K variants labelled *in vivo* with 32 P were removed from the SDS-polyacrylamide gel with a scalpel. The radioactive 70K proteins were removed from the gel fragments and eluted, and each was digested completely with trypsin. Equivalent quantities (*ca.* 10,000 cpm) of the tryptic peptides of the eleven 70K variants were then separated on thin-

Fig. 7. Tryptic phosphopeptide analysis of isoelectric 70K variants. The ^{32}P -labelled 70K variants were eluted from the gel shown in Figure 6 and digested totally with trypsin. Equal quantities of the resulting peptides were analysed on thin-layer cellulose plates by electrophoresis in the first and chromatography in the second dimension. In addition, a total-70K sample (containing all the 70K variants, without prior separation) was digested with trypsin and the products separated in the same way. The numbers in the panels refer to the variants (Figures 2–6) and Σ to the total-70K sample. The phosphopeptides were identified by autoradiography.

layer cellulose plates, with electrophoresis in the first dimension and chromatography in the second. The positions of the ^{32}P -labelled phosphopeptides in the chromatogram were found by autoradiography.

Figure 7 shows first of all the phosphopeptide analysis of all the 70K variants: the 70K protein has here been separated from the other snRNP proteins by elution from a one-dimensional SDS-polyacrylamide gel without prior IEF and then subjected to total tryptic digestion. While some spots can be assigned unambiguously to distinct phosphopeptides, for example phosphopeptides a, g and h, the phosphopeptides in the middle of the chromatogram are not so clearly separated. Short exposure of the autoradiograms revealed at least three distinct phosphopeptides, in the regions designated b, c, d and e. Only in the case of the phosphopeptide that remains close to the point of application on the chromatogram is this not entirely

unambiguous. The appearance and the position of this phosphopeptide varies from one chromatogram to another. It is therefore an open question whether we are seeing a single phosphopeptide that because of its low mobility, both in electrophoresis and in chromatography, appears diffuse, or whether we are seeing several different phosphopeptides. We regard it provisionally as a single phosphopeptide (f). We can thus distinguish at least eight different phosphopeptides in the set of 70K variants.

Panels 1–11 of Figure 7 show the phosphopeptide pattern of the 70K variants 1–11, corresponding to the labelling in Figure 6. A comparison of the phosphopeptide patterns of all variants shows both common features and differences. The phosphopeptides labelled most strongly, a, b, d and e, are found in all 70K variants. In contrast, phosphopeptides c, f, g and h are distributed differently and unevenly. As described above, the

position and the appearance of the f phosphopeptide vary markedly from one variant to another. Phosphopeptide c is seen in significant quantities only in the 70K variants 7–9, although it is frequently observable as a faint tail of phosphopeptide b (for example, in variants 1–5). The appearance of phosphopeptides a to h in the eleven 70K variants analysed is summarised in Table 2.

The two-dimensional separation of the 70K protein, as shown in Figure 6, is not sufficiently rigorous to prevent the possibility of cross-contamination of the neighbouring variant 'pairs' 3–4, 5–6 and 7–8, when these are excised from the gel. However, it is precisely these pairs that show significant differences in respect of the presence of the phosphopeptides f, g and h, respectively (Figure 7, Table 2); this confirms the correctness of the treatment of these protein spots as distinct 70K variants.

Surprisingly, in these 70K protein pairs it is always the variant with the higher pI value (the more basic variant) that shows one more phosphopeptide than the corresponding, more acidic variant. The same is found when 70K variants lying further from one another are compared. With variants 10 and 11, no significant difference in phosphopeptide pattern can be seen, even though their pI values are clearly different. In view of these observations, it seems highly unlikely that the differential phosphorylation of the 70K protein is the decisive factor in determining the large number of isoelectric variants that are separated by two-dimensional electrophoresis.

DISCUSSION

Characterisation of snRNP proteins by 2-D gel electrophoresis

This report describes the isoelectric characteristics of the U snRNP proteins, the occurrence of isoelectric variants of snRNP proteins and finally the evolutionary conservation of modification patterns of certain snRNP proteins between mouse and man.

Most of the snRNP proteins exhibit high pI values: some of them are even as basic as histones, with pI values between 10 and 10.6 (proteins A, B', B, B'', C, D and E). The basicity of the Sm antigenic proteins B, B' and D has already been suggested by immunoblot experiments with anti-Sm antisera, carried out on nuclear proteins separated by two-dimensional polyacrylamide gel electrophoresis (32–36).

The major exception among the snRNP proteins is F, which exhibits an acidic pI value of about 3.3. This is interesting for two reasons. It has been shown by Fisher *et al.* (29, 37) that in the absence of snRNAs the core proteins D, E, F and G associate in the cytoplasm to give a 6S complex. In view of the strongly basic character of proteins D, E and G, it is possible that by virtue of its acidic nature protein F may help to keep the 6S protein complex intact. This resembles to some extent the situation in the case of the histones. Here, the strongly basic histones H3 and H4 form a complex with the acidic nucleoplasmin (38). However, the analogy is limited. While nucleoplasmin promotes the integration of the histones H3 and H4 in nucleosomes without itself becoming incorporated into them, protein F after association of the 6S protein complex with newly-transcribed snRNA becomes a constitutive component of the snRNP core complex. In fact, protein F appears to be associated directly with regions of the snRNA in the core RNP, as was recently evidenced by the finding of cross-linking between protein F and snRNA induced by UV light (39).

The common proteins B, B', E, F and G usually migrate as discrete spots upon separation in two dimensions. This is

Table 2

Summary of the occurrence of tryptic phosphopeptide spots in the variants of protein 70K.

Variant	pI (from IEF)	Observed phosphopeptides	Variant pH
1	6.77	a b (c) d e (f) (g) h	6.5
2	6.88	a b (c) d e (f) (g) h	
3	7.08	a b (c) d e g h	7.0
4	7.12	a b (c) d e f (g) h	
5	7.33	a b (c) d e g (h)	7.5
6	7.44	a b (c) d e g h	
7	7.77	a b c d e f (g)	8.0
8	7.82	a b c d e f g (h)	
9	8.05	a b c d e f g (h)	8.5
10	8.18	a b (c) d e f (g)	
11	8.34	a b (c) d e f g	

Table 2. Distribution of the tryptic phosphopeptides in the 70K variants. The occurrence and intensity of the phosphopeptide spots detected in Figure 7 are shown along with the pI values of the 70K variants. Parenthesis around a letter indicate that the spot in question occurs with a much lower intensity in this peptide compared with in others. On the right the relative arrangement of the pI variants in a linear pH gradient is illustrated.

important, since it eliminates the possibility that each of the snRNPs U1, U2, U4/U6 and U5 might bear a slightly different variant of one or more of these common polypeptides—this with differences between the variants that are too small to allow detection in one-dimensional SDS-polyacrylamide gels.

For the protein D, the situation is not quite clear. The main spot, with a pI value of 10.5, is rarely symmetrical and frequently diffuse, with a downward tail, as though the protein were inhomogeneous. In addition, a further spot is always seen, with a pI value of 8.0 and a size equal to that of D, although this always occurs in disproportionately low amounts compared with D. In one-dimensional SDS-polyacrylamide gel separations of snRNP proteins, we have frequently observed a second band just ahead of D (28), which we have denoted D'. We therefore assign tentatively the protein with pI 8.0 to D'. We have recently obtained from the separation by HPLC of snRNP protein mixtures evidence for the existence of a third protein, present in all snRNPs and having a molecular weight of about 16 kDa (T. Lehmeier, K. Foulaki and R. Lührmann, unpublished). It is thus possible that the basic spot D contains at least two proteins that, because of their identical physico-chemical properties, cannot be separated from each other. The number and nature of the proteins with molecular weight ≈ 16 kDa will be resolved by protein-chemical analysis of the products of the HPLC separation of the D-sized proteins.

The grouping of the snRNP proteins into two classes, common and particle-specific, allows the assumption that the snRNP-specific proteins take part in the functional activity of the snRNP particles. The modification of the U1-specific protein C does not influence the electric charge of the protein, but rather increases

the molecular weight. This has already been shown clearly by pulse-chase experiments (29, 37). In the two-dimensional separation, C frequently produces a long, diffuse protein spot in the second dimension, sometimes two distinguishable spots. The U2-specific protein A' separates reproducibly into isoelectric variants. The chemical nature of the modifications of A' and C is unknown; however, we can exclude the possibility of phosphorylation.

Variants of the U1-specific 70K protein

The most complex modification pattern is seen with the U1-specific 70K protein, which occurs in more than 11 isoelectric variants that separate in the pI range 6.7–8.7. Since the 70K protein is the only one that is phosphorylated *in vivo*, we made the working hypothesis that the many 70K variants arose as a consequence of differential phosphorylation of the 70K chain. However, although we were able to show that all 70K variants are phosphorylated *in vivo*, several observations fail to accord with the working hypothesis.

If the observed forms of the 70K protein were generated by differential phosphorylation only, then the incorporation of phosphate groups should lead to a stepwise decrease in pI value of the 70K molecule by *ca.* 0.3–0.4 units per phosphate group, and a regular 'necklace' pattern would be observed in a linear pH gradient. Examples of this behaviour include, for example, the phosphorylated protein S6 of the 40S ribosomal subunit (40), the heat-shock protein hsp 28 (41) and the translation initiation factor eIF-4F (42). In addition, the intensity of the phosphorylation signal is frequently inversely proportional to the molecular weight of the phosphorylation variant; that is, the more highly phosphorylated protein molecules are comparatively underrepresented. However, in the case of the 70K protein the situation appears quite different. First of all, we observe a clear correlation between the masses of the individual 70K variants (measured by the intensity of Coomassie blue staining) and the intensities of the ³²P phosphate signals in the autoradiogram. Secondly, the differences in pI values between neighbouring 70K variants in the 2-D gel electrophoresis are very irregular, varying between 0.04 pI units for closely neighbouring pairs (for example, variants 3/4, 5/6 and 7/8) and 0.33 pI units for other pairs (for example 8/9).

The phosphopeptide analysis of the various 70K variants indicates a very complex situation. First, all the 70K variants possess four phosphopeptides in common (a, b, d and e), which suggests that all variants are constitutively phosphorylated at four serine residues. In addition to this, there are at least three further serines that are differentially phosphorylated in the various 70K variants. A comparison of the degree of phosphorylation of these serines in neighbouring 70K variants, however, shows no correlation with the pH values. For example, the phosphopeptide analysis of the closely situated 70K variants 3 and 4 reveals that it is the more basic variant, 4, that possesses an additional phosphopeptide and not, as would be expected, *vice versa* (Figure 2, Table 2).

With this in mind we are forced to conclude that the principal origin of most of the isoelectric variants of the 70K is another post-translational modification, such as the acetylation of lysine residues. Further indication of multiple types of modification of the 70K protein has been provided by the separation of the 70K variants in a 7% SDS-polyacrylamide gel after the isoelectric focussing in the first dimension. In this way, it becomes clear that some neighbouring variants, such as 1/2, 3/4, 5/6 and in

particular 7/8 differ significantly not only in their charge but also in their apparent molecular weight (Figure 4).

Independently of the identification of the chemical nature of the various modification of the 70K proteins, it must be asked why the presence of so many 70K variants should be needed for the function of the U1 snRNPs in the cell. It is imaginable that the degree of phosphorylation depends upon the cell cycle, and that the distribution and activity of the U1 snRNPs in the cell must be regulated during the different stages of this cycle. A particularly interesting question is whether the various types of U1 snRNP that differ in respect of their 70K variant also differ in respect of their activity in splicing. It should be possible to develop a test for this, based upon the total reconstitution of snRNPs and an assay of splicing activity *in vitro*.

A comparison of the snRNP proteins from HeLa cells and mouse Ehrlich ascites tumour cells shows a remarkable correspondence between the electrophoretic behaviour of the proteins of these two organisms. This correspondence is seen not only in the pI values of the main snRNP proteins, but also in the patterns of the post-translationally modified variants of these. All the 70K variants found in HeLa are also encountered in mouse. The evolutionary conservation of the modification pattern of certain snRNP proteins such as 70K, A' and C in species as distant as mouse and man points to an important role of these protein variants either in the biogenesis or in the function of the respective RNP particles with which these variants are associated. On the other hand, the absence of protein B' from Ehrlich ascites tumour cell snRNPs (this paper) as well as from snRNPs of other mouse cells (43) would suggest that B' is not an essential protein for the functioning of UsnRNPs in the cell. This is not surprising when it is remembered that proteins B and B' from human cells differ only in their carboxy-terminal part, when a proline-rich motif is repeated once more, in B' (44, 45).

ACKNOWLEDGEMENTS

We would like to acknowledge the expert technical assistance of Irene Öchsner-Welpelo and to thank Verena Buckow for help in the preparation of the manuscript. This work was supported by grants from the Bundesministerium für Forschung und Technologie and from the Deutsche Forschungsgemeinschaft to R.L. A.W. was supported in part by a fellowship from the Friedrich-Ebert-Stiftung.

REFERENCES

1. Reddy, R. and Busch, H. (1988). In: Structure and function of major and minor small nuclear ribonucleoprotein particles. Birnstiel, M.L. (ed.), Springer-Verlag, Berlin/New York, pp. 1–37.
2. Maniatis, T. and Reed, R. (1987). *Nature* **325**, 673–678.
3. Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Krämer, A., Frendewey, D. and Keller, W. (1988). In: Structure and function of major and minor small nuclear ribonucleoprotein particles. Birnstiel, M.L. (ed.), Springer Verlag, Berlin/New York, pp. 115–154.
4. Green, M.R. (1986). *Annu. Rev. Genet.* **20**, 671–708.
5. Guthrie, C. and Patterson, B. (1988). *Annu. Rev. Genet.* **22**, 387–419.
6. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983). *Cell* **33**, 509–518.
7. Heinrichs, V., Bach, M., Winkelmann, G. and Lührmann, R. (1989). *Science* **247**, 69–72.
8. Lührmann, R. (1988). In: Structure and function of major and minor small nuclear ribonucleoprotein particles. Birnstiel, M.L. (ed.), Springer Verlag, Berlin/New York, pp. 71–99.
9. Bach, M., Winkelmann, G. and Lührmann, R. (1989). *Proc. Natl. Acad. Sci. USA* **86**, 6038–6042.

10. Lerner, M.R. and Steitz, J.A. (1979). *Proc. Natl. Acad. Sci. USA* **76**, 5495–5499.
11. Tan, E.M. (1989). *Adv. Immunol.* **44**, 93–152.
12. Pettersson, I., Hinterberger, M., Mimori, T., Gottlieb, E. and Steitz, J.A. (1984). *J. Biol. Chem.* **259**, 5907–5914.
13. Habets, W., Hoet, M., Bringmann, P., Lührmann, R. and van Venrooij, W. (1985). *EMBO J.* **4**, 1545–1550.
14. Reuter, R., Rothe, S., Habets, W., van Venrooij, W.J. and Lührmann, R. (1990). *Eur. J. Immunol.* **20**, 437–440.
15. van Venrooij, W.J. and Sillekens, P. (1989). *Clinical and Experimental Rheumatology* **7**, 635–645.
16. Reuter, R. and Lührmann, R. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 8689–8694.
17. Bringmann, P., Rinke, J., Appel, B., Reuter, R. and Lührmann, R. (1983). *EMBO J.* **2**, 1129–1135.
18. Bochnig, P., Reuter, R., Bringmann, P. and Lührmann, R. (1987). *Eur. J. Biochem.* **168**, 461–467.
19. O'Farrell, P.H. (1975). *J. Biol. Chem.* **250**, 4007–4021.
20. Klose, J. (1983). *Mod. Meth. Prot. Chem.- Review Articles*, Tscheche, H. (ed.), de Gruyter & Co., Berlin/New York.
21. Dunn, M.J. and Burghes, A.H.M. (1983). *Electrophoresis* **4**, 97–116.
22. Laemmli, U.K. (1970). *Nature* **227**, 680–685.
23. Towbin, H., Staehelin, T. and Gordon, J. (1979). *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
24. Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Philipson, L. (1986). *EMBO J.* **5**, 3209–3217.
25. Smith, d.E. and Fisher, P.A. (1984). *J. Cell. Biol.* **99**, 20–28.
26. Bister, K., Trachmann, C., Jansen, H.W., Schroeder, B. and Patschinski, T. (1987). *Oncogene* **1**, 97–109.
27. Zieve, G.W. and Penman, S. (1981). *J. Mol. Biol.* **145**, 501–523.
28. Bringmann, P. and Lührmann, R. (1986). *EMBO J.* **5**, 3509–3516.
29. Fisher, D.E., Conner, G.E., Reeves, W.H., Blobel, G. and Kunkel, H.G. (1983). *Proc. Natl. Acad. Sci. USA* **80**, 6356–6360.
30. Gerke, V. and Steitz, J.A. (1986). *Cell* **47**, 973–984.
31. Wooley, J.C., Zuckerberg, L.R. and Chung, S.-Y. (1983). *Proc. Natl. Acad. Sci. USA* **80**, 5208–5212.
32. Schrier, W.H., Reddy, R. and Busch, H. (1982). *Cell Biol. Int. Rep.* **6**, 925–932.
33. Guldner, H.H., Lakomek, H.J. and Bautz, F.A. (1983). *J. Immunol. Methods* **64**, 45–59.
34. Schrier, W.H., Feinbaum, R. and Okamura, T.B. (1985). *Anal. Biochem.* **147**, 38–48.
35. Habets, W.J., Berden, J.H.M., Hoch, S.O. and van Venrooij, W.J. (1985). *Eur. J. Immunol.* **15**, 992–997.
36. Elkon, K.B. and Jankowski, P.W. (1985). *J. Immunol.* **134**, 3819–3824.
37. Fisher, D.E., Conner, G.E., Reeves, W.H., Wisniewski, R. and Blobel, G. (1985). *Cell* **42**, 751–758.
38. Kleinschmidt, J.A. and Franke, W.W. (1982). *Cell* **29**, 799–809.
39. Woppmann, A., Rinke, J. and Lührmann, R. (1988). *Nucleic Acids Res.* **16**, 10985–11004.
40. Thomas, G., Siegmund, M. and Gordon, J. (1979). *Proc. Natl. Acad. Sci. USA* **76**, 3952–3956.
41. Welch, W.J. (1985). *J. Biol. Chem.* **260**, 3058–3062.
42. Duncan, R., Milburn, S.C. and Hershey, J.W.B. (1987). *J. Biol. Chem.* **262**, 380–388.
43. Feeney, R.J., Sauterer, R.A., Feeney, J.L. and Zieve, G.W. (1989). *J. Biol. Chem.* **264**, 5776–5886.
44. Reuter, R., Rothe, S. and Lührmann, R. (1987). *Nucleic Acids Res.* **15**, 4021–4034.
45. van Dam, A., Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R. and Cuypers, H.T. (1989). *EMBO J.* **8**, 3853–3860.