Cain, A. K., & Nester, E. W. (1973) *J. Bacteriol. 115*, 769–776.
Chambon, P. (1975) *Annu. Rev. Biochem. 44*, 613–638.
Cheung, S.-C., Kobayashi, G. S., Schlessinger, D., & Medoff, G. (1974) *J. Gen. Microbiol. 82*, 301–307.
Gebert, U., Wieland, Th., & Boehringer, H. (1967) *Justus Liebigs Ann. Chem. 705*, 227–237.
Goodwin, R. A., Jr., & Des Prez, R. M. (1973) *South. Med. J. 66*, 13–25.
Grippo, P., Caruso, A., Locorotondo, G., & Labella, T. (1976) *Cell Differ. 5*, 121–128.
Groves, W. E., Davis, F. C., & Sells, B. H. (1968) *Anal. Biochem. 22*, 195–210.
Iverius, P. H. (1971) *Biochem. J. 124*, 677–682.
Kumar, B. V., McMillian, R., Medoff, G., Schlessinger, D., & Kobayashi, G. (1977) *Biochim. Biophys. Acta 478*, 192–200.
Laemmli, U. K. (1970) *Nature (London) 277*, 680–685.
Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., & Rutter, W. J. (1970) *Science 170*, 447–449.
Mauck, J. C. (1977) *Biochemistry 16*, 793–797.
Pong, S.-S., & Loomis, W. F., Jr. (1973) *J. Biol. Chem. 248*, 3933–3939.
Renart, J., & Sebastian, J. (1976) *Cell Differ. 5*, 97–107.
Roeder, R. G. (1974) *J. Biol. Chem. 249*, 249–256.
Roeder, R. G. (1976) *RNA Polymerases* (Losick & Chamberlin, Eds.) pp 285–329, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Roeder, R. G., & Rutter, W. J. (1969) *Nature (London) 224*, 234–237.
Schultz, L. S., & Hall, B. D. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1029–1033.
Shapiro, H. L., Vinnela, E., & Maizel, J. V. (1967) *Biochem. Biophys. Res. Commun. 28*, 815–819.
Sklar, V. E. F., & Roeder, R. G. (1975) *J. Biol. Chem. 250*, 1064–1073.
Virgil, E., Sklar, F., Yamamoto, M., & Roeder, R. G. (1976) *RNA Polymerases* (Losick & Chamberlin, Eds.) pp 803–817, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Young, H. A., & Whiteley, H. R. (1975) *Exp. Cell Res. 91*, 216–222.

Characterization of Glycine-Rich Proteins from the Ribonucleoproteins Containing Heterogeneous Nuclear Ribonucleic Acid†

Jean-Paul Fuchs, Clément Judes, and Monique Jacob*

ABSTRACT: The salt-soluble 28 000–38 000-dalton proteins were isolated from ribonucleoproteins containing heterogeneous nuclear RNA and partially purified. They were glycine-rich proteins (22–26 mol/100) and contained a small amount of N^G-dimethylarginine. Their N-terminal amino acid was blocked. Their p*I* was basic, extending from 6.95 to 9.20. Some 40 different polypeptides were demonstrated by combining molecular weight and p*I* determinations. Comparison of peptidic maps and of peptide size after trypsin and thermolysin digestion indicated the presence of only four proteins. The pattern of distribution of p*I* showing series of discrete major and minor bands common to two or three polypeptides of different apparent molecular weight was also compatible with the existence of four proteins and in addition supported the idea that the multiplicity of polypeptides was due to extensive posttranslational modifications.

The nuclear ribonucleoproteins containing heterogeneous nuclear RNA (hnRNP)[1] are made of two classes of constituents, 30–50S monoparticles and other complexes much more heterogeneous in size (Stévenin et al., 1977). In contrast to heterogeneous complexes, the monoparticles accumulated at 30–50S upon a mild ribonuclease treatment of the hnRNP (Samarina et al., 1968; Stévenin et al., 1970, 1977), they were easily dissociated by salt treatment (Gallinaro et al., 1975), and they were relatively poor in phosphoproteins (Stévenin et al., 1977; Gallinaro et al., 1975; Fuchs & Jacob, 1979). Approximately 20 major proteins could be detected in monoparticles by monodimensional gel electrophoresis (Stévenin et al., 1979). Some of them were acidic (p*I* of 5–7) and their molecular weight extended from 40 000 to 110 000. Another group was made of basic proteins (p*I* of 7–9), clustered between 28 000 and 38 000 daltons. They represented approximately 25% of the total hnRNP proteins. We started a study of these proteins in rat brain, our aim being, at long term, to determine whether the assembly of these basic hnRNP proteins may govern the structure of hnRNP as other basic proteins do in the assembly of chromatin or ribosomes.

A large-scale method had been devised previously, allowing the clear-cut separation of heterogeneous complexes from monoparticle proteins after salt dissociation (Fuchs & Jacob, 1979). The 28 000–38 000-dalton proteins could then be separated from the higher molecular weight proteins by gel chromatography. In the present work, we further fractionated the mixture of basic proteins by combining gel and carboxymethylcellulose chromatographies. The amino acid composition and the p*I* of individual proteins or groups of proteins were determined, and peptidic maps were established. The results support the idea of the existence of only four poly-

† From the Laboratoire de Génétique Moléculaire des Eucaryotes (CNRS), U184 (INSERM), Faculté de Médecine, 67000 Strasbourg, France. *Received July 30, 1979.* This work was partially supported by Grant ATP No. 2882 from the CNRS.

[1] Abbreviations used: RNP, ribonucleoproteins; hnRNP, ribonucleoproteins containing heterogeneous nuclear RNA.

0006-2960/80/0419-1087$01.00/0

peptidic chains extensively modified, so as to give rise to multiple polypeptides.

Materials and Methods

Fractionation of hnRNP Constituents. The method was described previously (Fuchs & Jacob, 1979). A nuclear extract containing hnRNP plus soluble nuclear proteins was prepared from purified brain nuclei (Stévenin & Jacob, 1974). The nuclear extract was chromatographed on DEAE-cellulose. The effluent containing the soluble nuclear proteins was discarded. The salt-soluble proteins were then eluted with 0.4 M NaCl together with some salt-resistant RNP. The eluate was adjusted to 1 M NaCl and centrifuged for 2 h at 50000 rpm (176000*g*) in order to remove the largest salt-resistant RNP. The proteins of the supernatant were precipitated with $(NH_4)_2SO_4$. The pellet was resuspended, dialyzed, and chromatographed on Bio-Gel P-150 (Bio-Rad Laboratories). The residual RNP were excluded and the material of 28000–38000 daltons was collected.

Chromatography on Carboxymethylcellulose. CM-cellulose (CM-52, Whatman) was precycled according to the instructions of the manufacturer and equilibrated in 250 mM sodium acetate, pH 5.5, 1 mM dithiothreitol, and 6 M urea. The CM-cellulose was packed in a glass column (6-mm diameter; 450-mm height) at 70 mL cm^{-2} h^{-1}. After being packed, the column was reequilibrated with CM medium (25 mM sodium acetate, pH 5.5, 1 mM dithiothreitol, and 6 M urea) at 40 mL cm^{-2} h^{-1} up to constant pH and conductivity.

Proteins from pooled Bio-Gel fractions were concentrated by $(NH_4)_2SO_4$ precipitation. The samples (6–11 mg of proteins) were dialyzed against CM medium and loaded on the column at 25 mL cm^{-2} h^{-1}. After being washed, the proteins were eluted with a linear gradient of NaCl (260 mL of CM medium; 0–0.2 M NaCl) at the same flow rate. Fractions of 0.75 mL were collected.

Polyacrylamide Gel Electrophoresis of Proteins and Peptides. The method was described previously (Gallinaro et al., 1975; Fuchs & Jacob, 1979). Essentially, the proteins were dissociated overnight in 10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, and 6 M urea, treated with 1% sodium dodecyl sulfate at 65 °C, and analyzed in 9% acrylamide slab gels. Staining was with Coomassie Brilliant Blue R.

When peptides were analyzed, a 15–25% linear gradient of acrylamide was used.

Electrofocusing. After precipitation with 10% trichloroacetic acid, the samples were washed twice with the same solution and twice with acetone. Reduction of S–S groups and alkylation of –SH groups were carried out as follows. The samples were treated overnight at 4 °C in the presence of 1.5 mM dithiothreitol in phosphate buffer, pH 7.0. *N*-Ethylmaleimide was added at a final concentration of 6 mM. The mixture was incubated for 30 min at 25 °C. The proteins were precipitated and washed as described above and then dissolved in a small volume of 8 M deionized urea. Electrofocusing was performed in an LKB multiphor according to the instructions of the manufacturer (Karlsson et al., 1973). All solutions were made up in 8 M deionized urea.

Amino Acid Composition. The proteins were hydrolyzed in doubly distilled 6 N HCl for 24 h at 108 °C. Amino acid compositions were obtained by using a Technicon automatic amino acid analyzer (Spackman et al., 1958).

Micromethod for Determination of N-Terminal Amino Acids. The method was described previously (Fuchs et al., unpublished experiments). Briefly, the proteins were precipitated with 10% trichloroacetic acid, washed with the same solution and with acetone, and dried under vacuum. Approximately 0.5 nmol of proteins was dissolved in 10 μL of H_2O. A total of 12.5 mg of urea, 7.5 μL of 0.2 M phosphate buffer, pH 8.85, and 12.5 μL of dimethylformamide was added in that order and left overnight at 4 °C. [^{14}C]Dansyl chloride (36 μCi) (CEA, Saclay, France; 100 mCi/mmol) in 10 μL of acetone was added. The tube was stoppered and left for 2 h in the dark with occasional shaking. The proteins were precipitated with trichloroacetic acid and electrophoresed in 10% acrylamide gel in the presence of 0.1% sodium dodecyl sulfate. The proteins were detected by fluorescence. The bands were excised and proteins were eluted out in the presence of 0.1% sodium dodecyl sulfate. The proteins were reprecipitated and hydrolyzed for 20 h at 115 °C in the presence of 6 N doubly distilled HCl. Two-dimensional chromatography was performed in glass jars according to Zanetta et al. (1970). The plates were autoradiographed for 1–3 days with Kodirex films (13 × 18 cm; Kodak).

Characterization of Dansyl-N^G-dimethylarginine. The proteins were hydrolyzed in 6 N HCl for 20 h at 115 °C. After being dried under vacuum, the amino acids were dansylated and separated by two-dimensional chromatography according to Zanetta et al. (1970). The dansylated amino acids were detected by fluorescence.

Peptidic Maps. The method which allows the detection of tyrosine-containing peptides was a modification of that of Elder et al. (1977). Fractions between 28000 and 38000 daltons collected after Bio-Gel chromatography were electrophoresed on slab gels in the presence of sodium dodecyl sulfate as described above. The plates were extensively destained in order to also maximally remove dodecyl sulfate. Destaining medium was 7.5% acetic acid–5% methanol. Each stained band was excised, transferred into a siliconed tube, and dried under vacuum.

A total of 40 μL of 0.5 M phosphate buffer, pH 7.5, 1 mCi of ^{125}I in 5 μL ($Na^{125}I$ without carrier; CEA, Saclay, France), and 10 μL of 10 mg/mL chloramine T in phosphate buffer was added to each tube in that order. After 30 min, the reaction was stopped by addition of 2 mL of 1 mg/mL sodium bisulfite. The unreacted ^{125}I was removed by several washings (300 mL each, per band) under constant stirring. The bands were cut into small fragments with a razor blade, transferred in a siliconed tube, and dried under vacuum.

NH_4CO_3 (50 mM, 1 mL), pH 7.8, containing 11–12 units either of trypsin (Worthington; 50 μg/mL) or of thermolysin (Merck; 25 μg/mL) was added to each sample. After incubation for 24 h at 37 °C under constant shaking, the supernatant was carefully pipetted off, centrifuged at low speed to remove residual gel fragments, and lyophilized. The residue was suspended in 0.5 mL of doubly distilled water and lyophilized again in order to remove residual bicarbonate. Peptides were then solubilized in 10 μL of acetic acid–formic acid–water (3:1:16 v/v/v).

Peptide fractionation was performed on cellulose-coated TLC plates (10 × 10 cm; Merck). A mixture of 2% (w/v) Orange G and 1% (w/v) acid fuchsin was used as the migration marker. First, the samples were electrophoresed in acetic acid–formic acid–water (3:1:16 v/v/v) in an LKB multiphor at 1 kV. Electrophoresis was stopped after about 15 min and the plate was air-dried. The second dimension was ascending chromatography in butanol–pyridine–acetic acid–water (32.5:25:5:20 v/v/v/v) and was performed in glass jars. The plates were air-dried again and autoradiographed for 2–24 h with Kodirex films.

Results

Fractionation of the Basic hnRNP Proteins on CM-cellu-

lose. When monoparticle proteins were analyzed by gel electrophoresis in the presence of sodium dodecyl sulfate, 10 well-individualized bands could be detected in the 28 000–38 000-dalton range. According to our nomenclature (Stévenin et al., 1978) which takes into account the existence of other nuclear proteins in that region, they were designated as 3a2, 3a4, 3b2, 3b3, 3c1, 3c2, 3d1, 3d2, 3e1, and 3e2 [Figure 2, panel II (D)]. The first number indicates the 28 000–38 000-dalton zone, the letter indicates a subzone (a group of bands of very close molecular weight), and the last number indicates individual bands within the subzone. 3c2 was the most prominent polypeptide in brain hnRNP. As the proteins studied here were all found in zone 3, the first number (3) will be omitted below.

After Bio-Gel chromatography, all 10 bands were found in a large peak (peak B, Figure 1, panel I). When the fractions were pooled and chromatographed on CM-cellulose, the proteins were eluted in the following order: c1, c2, b3–e1–e2, a4–a2–b2, d1, and d2. However, much overlapping was observed, and this was due to the presence of many polypeptides of close molecular weight and p*I* as will be shown below.

In order to partially palliate overlapping due to molecular weight, the material of the B Bio-Gel peak was pooled as indicated in Figure 1, panel I. Therefore, such groups of proteins as a2–a4–b2–b3 and d1–d2 which had close chromatographic behaviors on CM-cellulose (i.e., close p*I*) but different molecular weights could be separated beforehand (Figure 1, panel II). Each fraction was then chromatographed on CM-cellulose. Figure 2 (panel I) shows the elution profile of Bio-Gel fraction B3. The first large peak was rich in proteins c1 and c2. The relative amount of c1 vs. c2 decreased along the salt gradient (Figure 3) and only one fraction was made of electrophoretically homogeneous c2. Essentially the same elution profiles were obtained for Bio-Gel fractions B1 and B2 but with different relative proportions of the first peak. By pooling the fractions of the trailing part of the profile as indicated in Figure 2, panel I, we could isolate a2–a4–b2–b3 and d1–d2 without cross contamination from Bio-Gel fractions B1 and B3, respectively [Figure 2, panel II (A and C)]. The polypeptides c1 and c2 could be isolated from both fractions B2 and B3 [Figure 2, panel II (B)]. The recoveries were 23, 48, and 43% for a2–a4–b2–b3, c1–c2, and d1–d2, respectively. The polypeptides e1–e2 whose amount was low were found only as minor bands eluting together with a mixture of c2, b3, and a4 right after the large peak. Therefore, they could not be isolated by this procedure.

Amino Acid Composition. The amino acid composition was first determined on a–b, c1–c2, and d1–d2 (Table I). As expected from previous work on this group of proteins or on individual proteins of this group (Krichevskaya & Georgiev, 1969; Sarasin, 1969; Ishikawa et al., 1970; Martin et al., 1973; Christensen et al., 1977; Karn et al., 1977; Patel et al., 1978), the 28 000–38 000-dalton monoparticle proteins were characterized by a high proportion of glycine and by the presence of N^G-dimethylarginine (further proof of the presence of this modified amino acid will be given below). As the 28 000–38 000-dalton proteins were basic, a high proportion of basic vs. acidic amino acids was expected. This was not the case and we assume that a large proportion of Asx and Glx are Asn and Gln.

The composition of the basic monoparticle proteins was definitely different from that of two groups of acidic proteins of average molecular weight 60 000 and 50 000 (Table I, 2b–c and 2d–e). The latter were not as glycine-rich as the basic proteins but contained N^G-dimethylarginine (and possibly

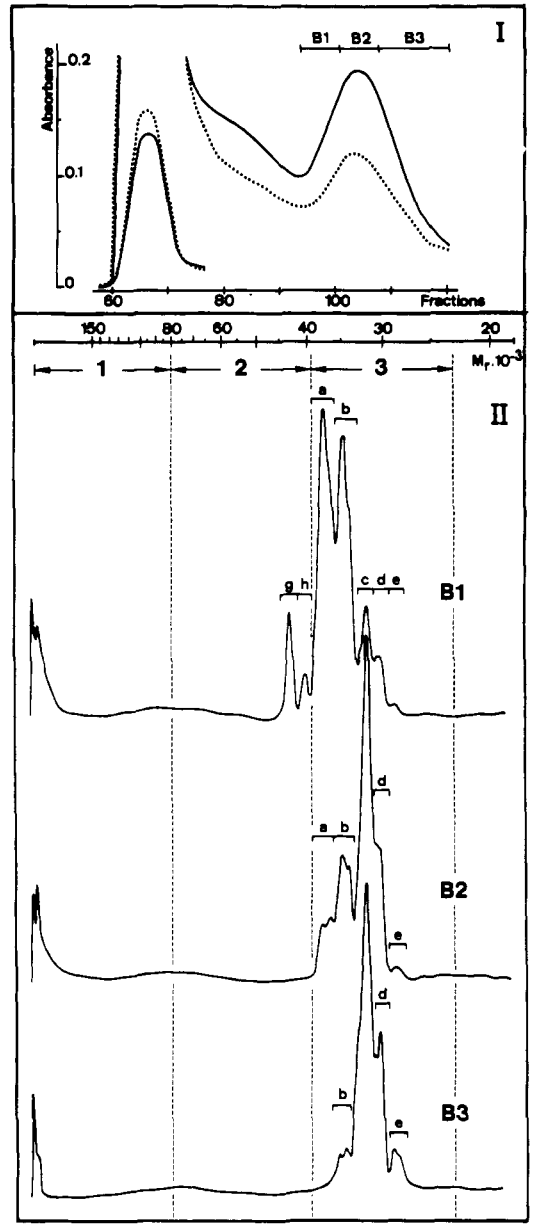

FIGURE 1: Bio-Gel P-150 chromatography. (Panel I) A fraction containing the salt-soluble monoparticle proteins plus ribonucleoproteins from heterogeneous complexes was chromatographed on Bio-Gel P-150 as previously described (Fuchs & Jacob, 1979). The void volume (up to fraction 72) contained the ribonucleoproteins. The 28 000–38 000-dalton monoparticle proteins were found in the last large peak (peak B, fractions 93–120). Fractions were pooled as indicated (B1, B2, and B3). (···) Absorbance at 260 nm; (—) absorbance at 280 nm. A 10-fold reduced scale is presented for the RNP peak. (Panel II) $NaDodSO_4$–polyacrylamide gel electrophoresis of fractions B1, B2, and B3.

traces of another modified amino acid). This indicated that methylation of arginine was not restricted to the basic proteins.

In spite of their striking similarities, the three groups (a–b, c, and d) of proteins were nevertheless different (Table I, underlined amino acids). This suggested a variability within the population and excluded that the multiple bands could be due solely to differences of conformation.

In order to verify whether such differences could also be observed between two polypeptides of very close molecular weights and close p*I* (as attested by chromatographic behavior), we analyzed the CM-cellulose chromatography fractions depicted in Figure 3. Fraction A contained approximately equal amounts of c1 and c2, the relative amount of c1 de-

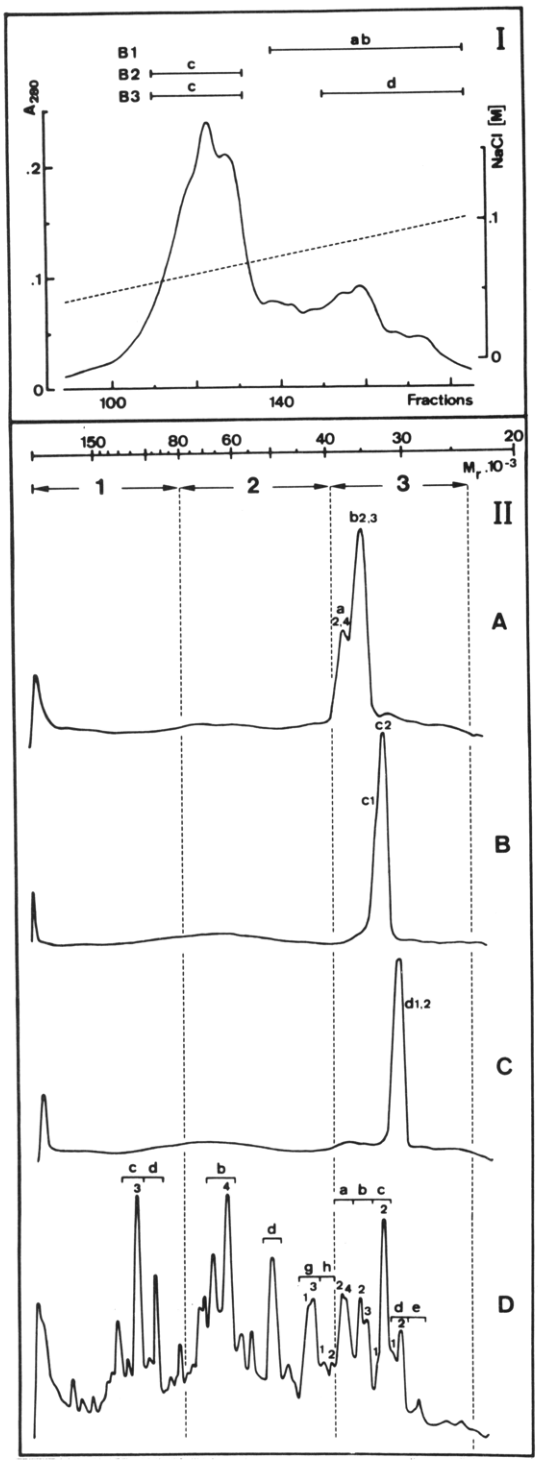

FIGURE 2: Carboxymethylcellulose chromatography. (Panel I) Bio-Gel fractions B1, B2, and B3 were chromatographed on CM-cellulose as described under Material and Methods. The proteins from the three fractions were eluted in the same NaCl concentration range. The elution profile shown here is that of B3. Those of B1 and B2 were close but the relative importance of the first large peak varied according to their content in c1–c2. The strategy of pooling is indicated for the three fractions. (Panel II) $NaDodSO_4$–polyacrylamide gel electrophoresis of fraction ab (from B1), c (from B2), and d (from B3). The profile of the proteins from a nuclear extract containing all the hnRNP proteins is given in (D).

creased up to fraction E which was electrophoretically pure c2. If the amino acid composition of c1 would be as different from c2 as, for instance, c is from d (Table I), a gradual variation of the amount of certain amino acids would be observed. This was not the case (Table II) and the compositions of the five fractions were virtually identical. Only a slight

Table I[a]

amino acid	amino acid composition				
	3ab	3c	3d	2b–c	2d–e
Asx	10.3	11.2	10.6	10.4	10.5
Thr	4.9	4.3	4.0	5.2	4.9
Ser	5.5	7.4	9.0	6.4	6.5
Glx	9.7	9.1	8.8	13.0	15.5
Pro	~5.5	3.8	3.3	6.1	5.7
Gly	25.7	25.3	22.4	12.5	8.2
Ala	3.3	3.5	4.9	6.2	7.1
Val	3.9	4.0	4.7	4.3	5.2
Cys	traces	traces	traces	0.8	0.7
Met	1.7	1.9	1.5	2.2	1.9
Ile	1.7	2.2	2.1	4.6	4.3
Leu	2.8	2.5	3.0	6.7	8.4
Tyr	5.3	5.8	4.0	3.5	2.6
Phe	4.9	5.5	6.1	3.7	3.6
Lys	5.4	4.9	5.9	6.3	6.4
His	1.9	2.0	2.5	1.5	1.8
$(CH_3)_2$Arg	1.0	0.5	1.2	0.7	0.7
Arg	6.2	6.2	6.0	5.9	6.0
Asx + Glx	20.0	20.3	19.4	23.4	26.0
Lys + His + Arg	13.5	13.1	14.4	13.7	14.2
(Asx + Glx)/ (Lys + His + Arg)	1.48	1.55	1.35	1.71	1.83

[a] Amino acid composition of isolated 28 000–38 000-dalton monoparticle proteins (mol/100 mol). 3a–b, 3c, and 3d are the isolated groups of proteins shown in Figure 2. The amino acids whose proportion varied significantly from one group to the other are underlined (two series of concordant experiments). 2b–c and 2d–e are two groups of acidic monoparticle proteins of average molecular weight 60 000 and 50 000 (Stephan, 1978).

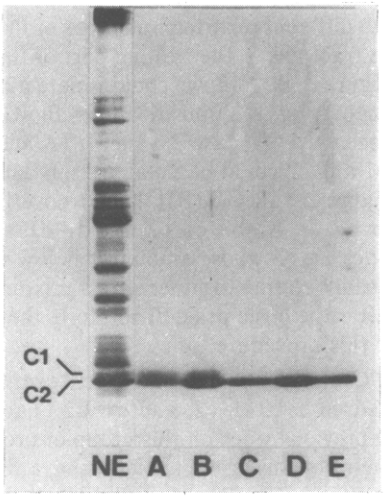

FIGURE 3: Electrophoretic analysis of proteins c. CM-cellulose chromatography was performed as described in Figure 2. The starting material was Bio-Gel fraction B2. Fractions were pooled in the portion of the salt gradient corresponding to the large peak of Figure 1, panel I (A–E). They are presented in the order of elution and were used for determination of amino acid composition (Table II) and p*I* determinations (Figure 6). NE is nuclear extract.

variation of the amount of Thr was observed whose significance remains to be determined. Therefore, with the possible exception of Thr, c1 and c2 had the same amino acid composition. As other proteins of close molecular weight were not separated, we ignore whether their composition would also be similar. Nevertheless, the results suggest that the population of glycine-rich polypeptides might be divided into a certain number of closely related proteins.

Table II[a]

amino acid	amino acid composition				
	A	B	C	D	E
Asx	11.3	11.7	12.6	12.0	11.1
Thr	4.9	4.7	4.0	3.3	3.2
Ser	7.3	7.9	8.1	7.9	7.3
Glx	8.8	8.9	8.9	8.8	9.1
Pro	~5.4	3.6	3.4	4.0	4.0
Gly	24.7	25.2	25.6	24.8	25.9
Ala	3.8	3.6	3.3	4.3	4.1
Val	3.7	3.8	3.7	3.9	3.8
Cys	traces	traces	traces	traces	traces
Met	1.8	1.9	1.8	1.9	1.8
Ile	2.2	2.1	2.0	2.2	2.0
Leu	2.3	2.3	2.3	2.4	2.2
Tyr	5.4	5.7	5.8	5.7	6.0
Phe	5.2	5.5	5.3	5.4	5.5
Lys	4.8	4.7	4.9	4.7	5.2
His	1.9	2.0	1.9	2.1	2.0
$(CH_3)_2$Arg	0.6	0.5	0.5	0.4	0.6
Arg	5.9	5.9	5.9	6.2	6.2
Asx + Glx	20.1	20.6	21.5	20.8	20.2
Lys + His + Arg	12.6	12.6	12.7	13.0	13.4
(Asx + Glx)/ (Lys + His + Arg)	1.60	1.63	1.69	1.60	1.51

[a] Amino acid composition of the 32 000–34 000-dalton monoparticle proteins (mol/100 mol). A–E are as defined in Figure 3.

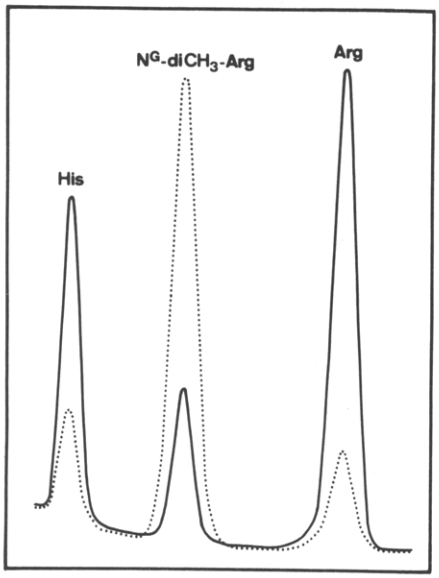

FIGURE 4: Demonstration of the presence of dimethylarginine in the hnRNP monoparticle proteins. An acid hydrolysate of proteins d (—) was chromatographed according to Spackman et al. (1958). A mixture of the usual amino acids plus $(CH_3)_2$Arg was chromatographed on an identical column (···). Only the pertinent portions of the profiles are presented.

Demonstration of the Presence of N^G-Dimethylarginine. When the acid hydrolysate was analyzed according to Spackman et al. (1958), an unknown peak of material was found between Arg and His (Figure 4). It was then shown that it cochromatographed with the marker N^G-dimethylarginine, suggesting the presence of this amino acid in the hydrolysate.

For verification, another procedure was adopted. After acid hydrolysis, the amino acids of the glycine-rich polypeptides were dansylated and the mixture was analyzed by thin-layer chromatography. In addition to the usual dansylated amino acids, another spot was detected (Figure 5, arrow) with the same chromatographic properties as dansyl-N^G-dimethylarginine.

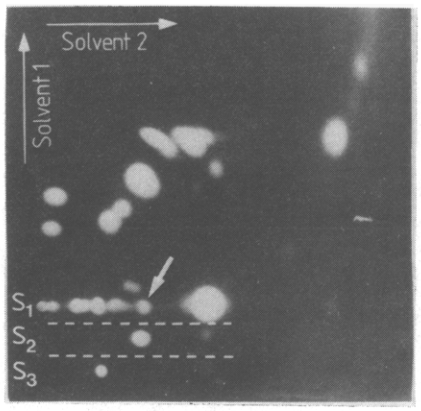

FIGURE 5: Demonstration of the presence of dimethylarginine in the hnRNP monoparticle proteins. An acid hydrolysate was dansylated and fractionated by thin-layer chromatography as indicated under Materials and Methods. S indicates start points. The acid hydrolysate of hnRNP protein c2 (S1), dansylated N^G-dimethylarginine (S2), and dansylated arginine (S3) served as markers. The arrow points to dansylated N^G-dimethylarginine in the protein hydrolysate.

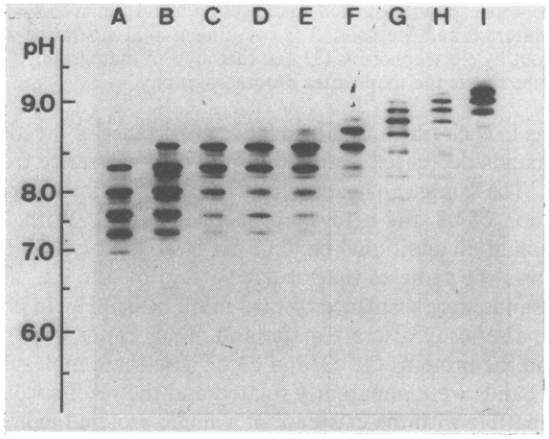

FIGURE 6: Electrofocusing of isolated proteins. Fractions collected after CM-cellulose chromatography were analyzed. (A–E) correspond to the fractions of group c depicted in Figure 3. (F) is enriched in b3 and c2, (G) and (H) contain a2–a4–b2 and (I) corresponds to d1–d2.

The two different methods applied to all the protein fractions confirmed that N^G-dimethylarginine was a normal constituent of the glycine-rich polypeptides and of other monoparticle proteins (Table I).

Determination of the N-Terminal Amino Acids. Under conditions where the N-terminal amino acid of known proteins could be unambiguously determined, no N-terminal amino acid was detected for any of the glycine-rich polypeptides. As will be seen below, several constituents of different p*I* were present in each of the isolated proteins or groups of proteins. The amount of proteins to be analyzed was such that a dansylated N-terminal amino acid would have been detected even if only one of the constituents had a free N-terminal amino acid.

We conclude (1) that the N-terminal amino acid of all the glycine-rich polypeptides is blocked and (2) that no proteolytic degradation which would have given rise to free N-terminal amino acids occurred during the isolation of the glycine-rich polypeptides.

Electrofocusing. Fractions containing purified or partially purified proteins eluted from CM-cellulose columns were analyzed. They are presented in the order of elution from CM-cellulose in Figure 6.

The first five fractions (A–E) correspond to the proteins of group c as shown in Figure 3. Several major and minor bands were detected between pH 6.95 and pH 8.69. Their p*I* was

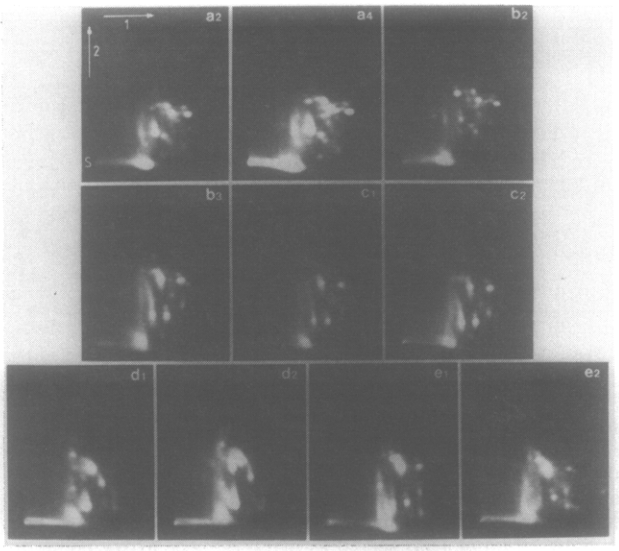

FIGURE 7: Tryptic maps. The 28 000–38 000-dalton proteins from Bio-Gel chromatography were separated by gel electrophoresis. The 10 bands were excised, iodinated, and digested by trypsin as indicated under Materials and Methods. The two-dimensional separation was carried out by electrophoresis (1) and then by chromatography (2). The figure shows the maps after autoradiography.

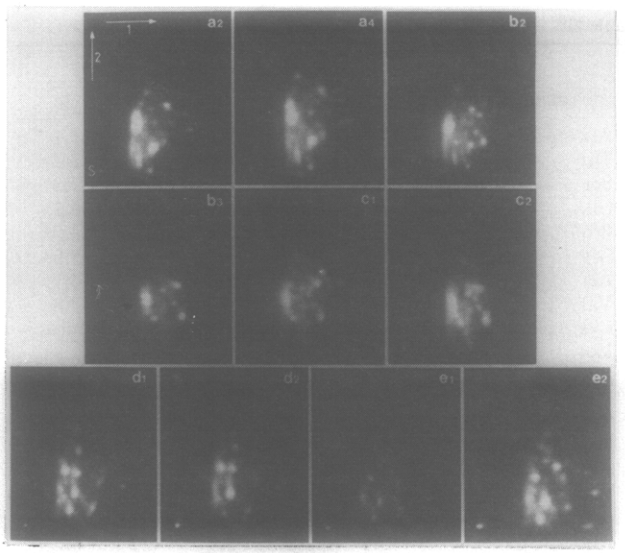

FIGURE 8: Thermolysin maps. The same material and the same method were used as described in Figure 7.

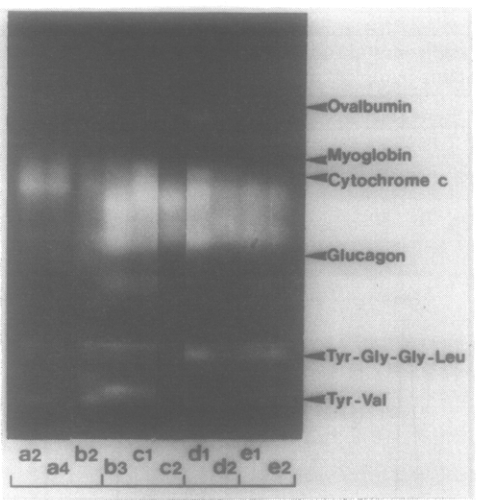

FIGURE 9: Size of the peptides formed upon tryptic digestion of hnRNP monoparticle proteins. Autoradiography after slab gel electrophoresis in a 15–25% linear acrylamide gradient. The same material was analyzed as described in Figure 7. The molecular weights of the markers were as follows: ovalbumin, 45 000; myoglobin, 18 000; cytochrome *c*, 12 400; glucagon, 3470; Tyr-Gly-Gly-Leu, 556, Tyr-Val, 280.

the same in each fraction. The relative importance of the most acidic bands decreased gradually up to disappearance from A to E. The sample analyzed in F primarily contained a large proportion of b3 and a lower proportion of c2. With the exception of an additional band at pH 8.80, the p*I* of all the bands was the same as that of bands from c1 and c2. The major bands were displaced toward more basic pH. In conclusion, a series of discrete major and minor bands were detected in the proteins c1, c2, and b3. These three molecular weight bands were apparently related and the results would be compatible with the existence of a single protein modified to various extents. This will be discussed.

The samples analyzed in G and H contained a2, a4, and b2, the relative proportion of b2 being higher in H than in G. The same bands were detected between pH 7.85 and pH 8.98 in the two samples. Additional minor bands were present at more acidic pH in G and more basic pH in H. The mobilities were different from those observed for the group c1, c2, and b3 with the exception of some minor bands. The presence of such bands might be due to contamination with a small amount of b3 or merely to chance. Though less fractions could be analyzed than for the group c1, c2, and b3, the results show the same general pattern of several bands with gradual variations of intensity.

The proteins d1 and d2 were not separated. But in their case also, several bands of different p*I* were detected. With one exception probably due to chance (at pH 8.98), the mobility of the bands was different from those of the two other groups.

The results indicate that each of the eight proteins separated according to molecular weight contains several p*I* bands. The distribution of the p*I* bands suggests the presence of three groups of proteins. Twenty bands were detected for c1–c2–b3, twelve were detected for a2–a4–b2, and eight were detected for d1–d2, i.e., 40 bands in total (not including e1 and e2).

Peptidic Maps and Size Distribution of Peptides. As the method used for determination of peptidic maps after iodination required only a low amount of material, the 10 bands separated by gel electrophoresis could be analyzed independently. In particular, this allowed the study of proteins e1 and e2 which could not be purified by CM-cellulose chromatography. Trypsin and thermolysin digestions were performed.

Tryptic maps are shown in Figure 7. Some of the peptides migrated in the first dimension but not in the second due to their insolubility in the chromatographic solvent. Four groups with identical fingerprints were apparent: (1) a2, a4, and b2; (2) b3, c1, and c2; (3) d1 and d2; (4) e1 and e2. Thermolysin maps are represented in Figure 8. The same four groups of proteins as described above were determined. Some differences in the relative intensities of certain spots were observed but their signification is not known. A careful comparison suggests that part of the thermolysin peptides could be common to groups a2–a4–b2 and b3–c1–c2, which would be in agreement with the results of amino acid composition. That this corresponds to the identity of sequences of a certain chain fragment has to be confirmed by other means.

The approximate size of the peptides was determined by electrophoresis (Figure 9). After tryptic digestion, peptides were particularly abundant in the regions of di- and tetrapeptides on one side and between 3500 and 20 000 daltons on the other. Peptides of size intermediary between these two

regions were only found in certain proteins. The group classification as proposed above (a2–a4–b2, b3–c1–c2, d1–d2, and e1–e2) was confirmed.

Other points are worth mentioning. (1) The proteins a2, a4, and b2 (35000–38000 daltons) gave rise to several peptides between 10000 and 20000 daltons. Several large-size peptides were also detected in the other proteins. A component with the approximate molecular weight of the undigested protein was even present in group d in addition to several peptides from 5000 to 20000 daltons. (2) Although the overall peptide distributions were close to each other within a group, the quantitative distribution of the peptides changed for each individual band. To interpret these facts, we propose that the proteins were modified at several sites (in particular Lys and Arg) to various extents. Tryptic digestion would thus produce series of peptides whose size would be related to the site of the modification.

The results were not as clear-cut after thermolysin digestion as most of the peptides were clustered at 3000–7000 daltons and could not be studied. These results will not be commented on further.

Discussion

Several characteristics were common to all the brain hnRNP salt-extractable proteins found in the 28000–38000-dalton range. The most striking one was the high content of glycine which varied from 22 to 26 mol %, corresponding to 66–91 residues per molecule. The presence of a small amount of methylated arginine (0.5–1.2 mol %; 1–4 residues per molecule) was also a constant feature of this group of proteins. Furthermore, the N-terminal amino acids of all the proteins were blocked. Another interesting characteristic was the low amount of basic relative to acidic amino acids in spite of a clearly basic p*I*. This might be due to the presence of glutamine and asparagine, an alternative being the introduction of basic groups or the blocking of the acidic groups by post-translational modifications. This latter possibility was suggested by the study of p*I*.

hnRNP proteins in the same molecular weight range were studied by several other groups (Krichevskaya & Georgiev, 1969; Sarasin, 1969; Ishikawa et al., 1970; Martin et al., 1973; Christensen et al., 1977; Karn et al., 1977; Patel et al., 1978). Whatever the cell type from which the hnRNP were extracted, the amino acid composition of the proteins was close to that found for brain hnRNP. In particular, the glycine content was always high as compared to that of other known cellular proteins. We therefore propose to designate the basic 28000–38000-dalton proteins from hnRNP as "glycine-rich proteins" (GRP).

A review of the literature showed that the number of proteins described between 25000 and 40000 daltons varied to a large extent according to the method of analysis. Monodimensional gel electrophoresis showed the presence of 1–10 proteins (Krichevskaya & Georgiev, 1969; Niessing & Sekeris, 1971; Albrecht & Van Zyl, 1973; Ducamp & Jeanteur, 1973; Gallinaro & Jacob, 1973; Martin et al., 1973; Pederson, 1974; Faiferman & Pogo, 1975; Beyer et al., 1977), but 15–21 proteins were detected by two-dimensional gel electrophoresis (Karn et al., 1977; Brunel & Lelay, 1979; Maundrell & Scherrer, 1979). The determination of the p*I* of isolated proteins shows a much larger number of bands, 8 molecular weight bands (out of 10) giving rise to 40 p*I* bands. It is likely that the successive determination of molecular weight and p*I* allows a better resolution than two-dimensional gel electrophoresis as minor bands whose p*I* and molecular weight are close to that of the major ones probably escape detection by the latter method.

Table III[a]

bands	M_r	p*I* range	proposed designation
a2	37700	7.25–9.05	GRP 1
a4	37000		
b2	35300		
b3	34500	6.95–8.80	GRP 2
c1	33600		
c2	32300		
d1	30800	7.65–9.20	GRP 3
d2			
e1	29600	n.d.[b]	GRP 4
e2	29100		

[a] Classification of the glycine-rich proteins. The tryptic peptide and thermolysin peptide maps, the tryptic peptide size distribution, and the pattern of electrofocusing bands suggest that the 10 molecular weight bands are modified forms of four unique polypeptides. [b] n.d., not determined.

However, our results suggest that there are not as many proteins of different primary structures but only four proteins modified to various extents. The quantitatively most important group was made of the three M_r bands designated as c1, c2, and b3. Peptidic maps after trypsin and thermolysin digestion were very close if not identical, which confirmed the similarity of amino acid compositions observed for partially separated c1 and c2. The characteristic pattern of p*I* bands with gradual shift from pH 7 to pH 8.8 suggested that if the same protein was present, modifications drastically changed its p*I*. It is not known whether the unmodified polypeptidic chain is acidic (presence of Asp and Glu) or basic (presence of Asn and Gln), but whatever its p*I*, the results can be explained only if most of the molecules were modified. The major p*I* bands are probably the consequence of such quantitatively important modifications. These polypeptides might, in turn, be partially modified at other sites, giving rise to the minor bands. The methylation of arginine is an example of such minor modifications. Phosphorylation which concerns only a fraction of the monoparticle proteins (Fuchs & Jacob, 1979) might be another example. The nature of the major and minor modifications is currently investigated. Periodic acid staining was positive, suggesting that glycine-rich proteins might well be glycoproteins.

Assuming that our interpretation of the results is correct, it is worth mentioning that the modifications are accompanied by changes of apparent molecular weights up to 2000 daltons. This might be due to the molecular weight of the additional constituents, to differences of dodecyl sulfate binding as a consequence of modifications, or to other yet undetermined causes. Changes of apparent molecular weight due to protein phosphorylation were described (Lamb & Choppin, 1977).

Due to the coexistence of many proteins of close p*I* and molecular weights, it is not surprising that clear-cut chromatographic separation according to charge and mass could not be achieved. The knowledge of the nature of the major modification might be of help for overcoming such difficulties in the future. It remains that the determination of several of the characteristics of the glycine-rich polypeptides gave a strong indication of the presence of only four polypeptidic chains extensively modified. Though their existence will be unambiguously established only after the direct demonstration of modifications and the determination of primary structures, the evidences obtained up to now seem to be sufficient to allow a classification of the glycine-rich proteins from hnRNP as shown in Table III.

Acknowledgments

We are very grateful to Dr. H. Duranton for making available to us the facilities for amino acid analysis and to Dr. D. Collot and M. F. Castaldi for performing the analyses. The expert technical assistance of L. Kister is acknowledged. We thank Dr. K. Maundrell and C. Brunel for making their results available to us before publication.

References

Albrecht, C., & Van Zyl, I. M. (1973) *Exp. Cell Res. 76*, 8–14.

Beyer, A. L., Christensen, M. E., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell 11*, 127–138.

Brunel, C., & Lelay, M. N. (1979) *Eur. J. Biochem. 99*, 273–314.

Christensen, M. E., Beyer, A. L., Walker, B. W., & LeStourgeon, W. M. (1977) *Biochem. Biophys. Res. Commun. 74*, 621–629.

Ducamp, C., & Jeanteur, Ph. (1973) *Biochimie 55*, 1235–1243.

Elder, J. H., Pickett, R. A., II, Hampton, J., & Lerner, R. A. (1977) *J. Biol. Chem. 252*, 6510–6515.

Faiferman, I., & Pogo, A. O. (1975) *Biochemistry 14*, 3808–3816.

Fuchs, J.-P., & Jacob, M. (1979) *Biochemistry 18*, 4202–4208.

Gallinaro, H., & Jacob, M. (1973) *FEBS Lett. 36*, 105–108.

Gallinaro, H., Stévenin, J., & Jacob, M. (1975) *Biochemistry 14*, 2547–2554.

Ishikawa, K., Kuroda, C., & Ogata, K. (1970) *Biochim. Biophys. Acta 213*, 505–512.

Karlsson, C., Davies, H., Ohman, J., & Andersson, U. B. (1973) Application Note L.K.B., Sweden, 1–13.

Karn, J., Vidali, G., Boffa, L. C., & Allfrey, V. G. (1977) *J. Biol. Chem. 252*, 7307–7322.

Krichevskaya, A. A., & Georgiev, G. P. (1969) *Biochim. Biophys. Acta 194*, 619–621.

Lamb, R. A., & Choppin, P. W. (1977) *Virology 81*, 382–397.

Martin, T., Billings, P., Levey, A., Ozarslan, S., Quinlan, T., Swift, H., & Urbas, L. (1973) *Cold Spring Harbor Symp. Quant. Biol. 38*, 921–932.

Maundrell, K., & Scherrer, K. (1979) *Eur. J. Biochem. 99*, 225–246.

Niessing, J., & Sekeris, C. E. (1971) *Biochim. Biophys. Acta 247*, 391–403.

Patel, N. T., Kurosky, A., & Holoubek, V. (1978) *Biochim. Biophys. Acta 533*, 282–286.

Pederson, T. (1974) *J. Mol. Biol. 83*, 163–183.

Samarina, O. P., Lukanidin, E. M., Molnar, J., & Georgiev, G. P. (1968) *J. Mol. Biol. 33*, 251–263.

Sarasin, A. (1969) *FEBS Lett. 4*, 327–330.

Spackman, D. H., Stein, W. H., & Moore, S. (1958) *Anal. Chem. 30*, 1190–1206.

Stéphan, D. (1978) Thesis, Université Louis Pasteur de Strasbourg, France.

Stévenin, J., & Jacob, M. (1974) *Eur. J. Biochem. 47*, 129–137.

Stévenin, J., Mandel, P., & Jacob, M. (1970) *Bull. Soc. Chim. Biol. 52*, 703–720.

Stévenin, J., Gallinaro, H., Gattoni, R., & Jacob, M. (1977) *Eur. J. Biochem. 74*, 589–602.

Stévenin, J., Gattoni, R., Gallinaro, H., & Jacob, M. (1978) *Eur. J. Biochem. 84*, 541–549.

Stévenin, J., Gattoni, R., Devilliers, G., & Jacob, M. (1979) *Eur. J. Biochem. 95*, 593–606.

Zanetta, J.-P., Vincendon, G., & Gombos, G. (1970) *J. Chromatogr. 51*, 441–458.