The EMBO Journal vol.8 no.4 pp.1177 - 1181, 1989

cDNA sequence of the rat U snRNP-associated protein N: description of a potential Sm epitope

George McAllister, Alison Roby-Shemkovitz, Susan G.Amara and Michael R.Lerner

Section of Molecular Neurobiology, Howard Hughes Medical Institute Research Laboratories, Yale University School of Medicine, 333 Cedar Street, PO Box 3333, New Haven, CT 06510, USA

Communicated by J.A.Steitz

Anti-Sm antibodies from a patient with systemic lupus erythematosus (SLE) were used to isolate cDNA clones encoding the snRNP-associated protein N from a rat brain derived cDNA library. The predicted primary structure of the 240 amino acid protein has a proline rich carboxyl terminus and shares a region of sequence similarity with other snRNP polypeptides, A and B/B'. Anti-Sm sera recognize a β-galactosidase fusion protein containing only the carboxyl-terminal 80 amino acids of N; antibodies eluted from this fusion protein also react with A, B/B' and N on immunoblots, suggesting that these proteins share an Sm epitope located within this segment. Polyclonal antibodies raised against a 23 amino acid synthetic peptide derived from this conserved region of N recognize A, N and B/B' on immunoblots and can immunoprecipitate the Sm class of U snRNAs. These results confirm that this sequence defines a potential Sm epitope. RNA blotting analyses demonstrate that a 1.6 kb mRNA expressed predominantly in brain encodes the N polypeptide in both rats and humans. At low stringency rat N cDNA also hybridizes to a 1.3 kb mRNA species which encodes B/B', suggesting that N is structurally related to, but distinct from B/B'. Although B/B' proteins are thought to be expressed in all human cells, only N and B, but not B', are observed on immunoblots of human brain proteins probed with anti-Sm sera. The apparent difference in the complement of proteins associated with snRNP particles in human brain versus elsewhere suggests a possible mechanism for the regulation of brain-specific mRNA splicing.

Key words: autoimmunity/cDNA cloning/small nuclear ribonucleoproteins/Sm antigen

Introduction

The Sm class of U small nuclear ribonucleoprotein particles (snRNPs) play a key role in the autoimmune disease systemic lupus erythematosus (SLE) (Lerner and Steitz, 1979; Tan, 1982) and are essential for the splicing of nuclear mRNA precursors (pre-mRNAs) (reviewed in Steitz *et al.*, 1988). In human cells, the Sm snRNPs contain at least seven common polypeptides, termed B', B, D, D', E, F and G; three U1-specific polypeptides, '70K' (70 kd), A and C; and two U2-specific polypeptides, A' and B'', complexed with the U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs). Anti-Sm sera recognize the polypeptides B/B', D, and to a lesser extent E. We have recently reported the existence of another Sm polypeptide, termed N, which is expressed in a tissue-specific manner in rats (McAllister *et al.*, 1988). The identification of the Sm antigenic epitope(s) is necessary to investigate its relevance to autoimmune disease. B and B' are thought to be closely related proteins, and may even be products of the same gene (Luhrmann, 1988). In rats, B' is either not expressed or co-migrates with B on gels, therefore, we will refer to these proteins as B/B' throughout.

Although precise mechanisms remain to be elucidated, it is now generally accepted that snRNPs are essential for the nuclear splicing of pre-mRNAs. For example, the U1 snRNP acts at the 5' splice site, whereas the U2 snRNP interacts with the branch point and the U5 snRNP probably acts at the 3' splice site (reviewed in Steitz *et al.*, 1988). However, a potential role for snRNPs in the mechanism(s) of tissue-specific or developmentally regulated alternative splicing has not been widely investigated. Determination of the regional distribution of snRNPs and their components may provide some insight into this problem. Likewise, the molecular cloning of snRNP proteins will be useful in understanding the structure and function of snRNP particles, and also provide information about the role of these proteins in autoimmune disease. Recently, the cDNA cloning of several snRNP polypeptides has been reported (Wieben *et al.*, 1985; Theissen *et al.*, 1986; Habets *et al.*, 1987; Sillekins *et al.*, 1987; Rokeach *et al.*, 1988; Yamamoto *et al.*, 1988). Here, we report the cDNA sequence of the rat snRNP polypeptide N, define a potential Sm epitope and investigate the complement of proteins associated with snRNPs in human brain. A possible mechanism for the regulation of tissue-specific pre-mRNA splicing is suggested.

Results

The isolation and partial characterization of cDNA clones encoding the tissue-specific, snRNP-associated polypeptide N has been previously described (McAllister *et al.*, 1988). Figure 1 shows a schematic diagram of these clones and the complete nucleotide sequence is given in Figure 2. No difference was observed in the overlapping sequence of the clones. There is one large open reading frame (ORF), between nucleotides 463 and 1185, capable of encoding a polypeptide of 240 amino acids. Three potential initiation codons occur at the beginning of this ORF, the first of which is probably the actual initiation codon because it diverges least from the consensus initiation sequence, ACCATGG (Kozak, 1986). There is a translational stop codon (TAA) at position 1183 and at the 3' end of the sequence there is a stretch of 16 adenines, preceded by the polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976).

The amino acid sequence of N has a predicted mol. wt of 24.614 kd, somewhat less than the 28 kd estimated from SDS-PAGE. However, when RNA transcripts made from

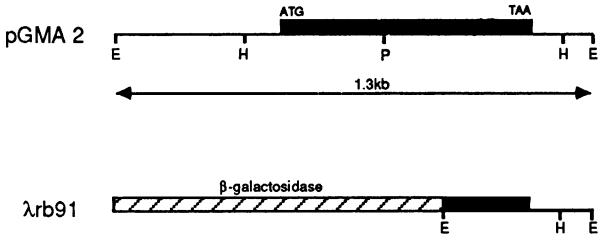

Fig. 1. A schematic diagram of cDNA clones encoding N. Clone λrb91 encodes a β-galactosidase−N fusion protein containing the carboxyl-terminal 80 amino acids of N. Clone pGMA2 contains the whole coding region of N as well as 5′ and 3′ flanking regions.

```
1     AGAGAGTAGAGGGGTGCTGGAGATGCCAGACGGTTGGTTCTGAGGAGAGATTTTGCAACGAATGGAGCGAGGAAGGGATCGTTTACACTT
91    GAGAAGAACTACTGAACAGCACGTGCCAGAGATTGAGGTCCAGGTCAAACGTAGAAGGACAGCCTCACTGAGCAACCAAGAGTGTCACTT
181   GTACCCACGGCGTTCTCAGCAACAGCAAATTCCTGTGGTGGATTTCCAGGCAGAACTGAGACAGGCGTTCTTAGCTGAGACACCAAGAGG
271   TGGTTAAAGCAGTATTGGAACTTCAAGGTGGTGGAAGTCAACAAACACAGGACCTATCCACTGATTGTGAAACTTTGGTCAAGCTTACAC
361   TGTGTTAATAGCCCTGCATCAAACCTTTATTTATTGCCCTTCTCCAAGTATTAAGGATCTTGAAATTTTAGTGTTGACAACTGCTATTGT
                MetThrValGlyLysSerSerLysMetLeuGlnHisIleAspTyrArgMetArgCysIleLeuGlnAspGlyArgIle      26
451   GGAACAGCAATCATGACTGTGGGTAAGAGTAGCAAGATGCTCCAGCATATTGACTATAGAATGAGATGTATCCTGCAAGATGGAAGAATC
      PheIleGlyThrPheLysAlaPheAspLysHisMetAsnLeuIleLeuCysAspCysAspGluPheArgLysIleLysProLysAsnAla      56
541   TTCATTGGCACCTTTAAGGCTTTTGACAAGCATATGAATTTGATCCTCTGTGATTGTGATGAGTTCAGGAAGATCAAGCCAAAGAATGCA
      LysGlnProGluArgGluGluLysArgValLeuGlyLeuValLeuLeuArgGlyGluAsnLeuValSerMetThrValGluGlyProPro      86
631   AAACAGCCAGAACGTGAAGAAAAACGGGTTTTGGGTCTGGTCTTGCTACGTGGAGAGAACTTGGTTTCCATGACAGTGGAGGGTCCACCT
      ProLysAspThrGlyIleAlaArgValProLeuAlaGlyAlaAlaGlyGlyProGlyValGlyArgAlaAlaGlyArgGlyValProAla     116
721   CCTAAAGATACTGGCATTGCTCGTGTGCCACTTGCTGGTGCTGCAGGTGGCCCTGGTGTTGGAAGACGAGCTGGCAGAGGAGTACCAGCA
      GlyValProIleProGlnAlaProAlaGlyLeuAlaGlyProValArgGlyValGlyGlyProSerGlnGlnValMetThrProGlnGly     146
811   GGTGTACCTATTCCCCAAGCTCCTGCTGGATTAGCAGGCCCTGTCCGAGGAGTTGGAGGCCCATCCCAGCAGGTCATGACCCCACAGGGA
      ArgGlyThrValAlaAlaAlaAlaAlaValAlaAlaThrAlaSerIleAlaGlyAlaProThrGlnTyrProProGlyArgGlyThrProPro     176
901   AGAGGCACTGTTGCAGCTGCTGCTGTTGCTGCTACTGCTAGCATTGCAGGAGCCCCAACCCAGTACCCGCCAGGACGGGGAACTCCACCT
      ProProValGlyArgAlaThrProProProGlyIleMetAlaProProProGlyMetArgProProMetGlyProProIleGlyLeuPro     206
991   CCACCTGTAGGCAGAGCAACCCCAACTTCCAGGCATTATGGCTCCTCCACCTGGAATGAGACCACCCATGGGCCCACCAATTGGACTTCCC
      ProAlaArgGlyThrProIleGlyMetProProProGlyMetArgProProProProGlyIleArgGlyProProProGlyMetArg     236
1081  CCTGCTCGAGGGACACCTATAGGCATGCCCCCTCCAGGAATGAGACCCCCTCCTCCAGGAATGAGAGCCCCACCTCCCCCAGGAATGCGT
      ProProArgProEnd                                                                              240
1171  CCACGAAGACCCTAAGATACAGTTGATAAATCTCAGCCCTTCTCTTTCCCTACAATGCTTCTTGTGAAATTGTGTCGCCTGCAAGCTTTT
1261  GACCCCTCTTACTGCATTAACTATAGATAATAAATACATAGCGCAATTGAAAAAAAAAAAAAAA  1325
```

Fig. 2. Nucleotide sequence of the rat N cDNA clone (pGMA2) and the predicted amino acid sequences of N. Nucleotides are numbered at left and amino acids at right. The translation initiation and termination codons and the poly(A) signal are underlined.

```
N  ATPPPG-IMAPPPGMRPPMGPPIGLPPARGTPIGMPPPGMRPPPPGIRGPPPPGMRPPRP 240
   |||  | : :| |||:|   :|  ::: :| |  |||||| |||   :|: ||| :|||:
A  ATPVVGAVQGPVPGMPPMTQAPRIMHHMPGQPPYMPPPGMIPPPGLAPGQIPPGGMPPQQ 189
```

Fig. 3. Amino acid sequence homology between the rat N and human A snRNP-associated proteins. The alignment was carried out according to the algorithm of Lipman and Pearson (1985). Vertical lines represent amino acid identity and dots represent conservative substitutions. A peptide was synthesized corresponding to residues 216−238 of N (represented by a line above the sequence), and used to raise antibodies in rabbits. The sequence of human A is as reported by Sillekens *et al.* (1987).

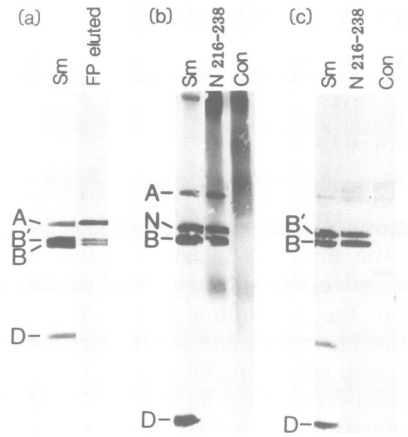

Fig. 4. Immunoblots of snRNP proteins. (**a**) Semi-purified U1 snRNPs from human HeLa cells probed with an anti-Sm serum (lane Sm), and with antibodies eluted from the λrb91 encoded fusion protein after an initial probing with anti-Sm serum (lane FP eluted). (**b**) snRNP proteins immunoprecipitated from rat MTC cells probed with anti-Sm serum (lane Sm), rabbit antibodies raised against residues 216−238 of N (Lane N216−238), and pre-immune rabbit serum (lane Con). (**c**) Total protein from HeLa cells probed with the same antibodies as in (b). The snRNP proteins A, N, B, B′ and D are indicated.

the full length cDNA clone are translated in an *in vitro* translation system the product also has an apparent mobility of 28 kd (McAllister *et al.*, 1988). Prolines make up 20% of the protein, and are concentrated mostly in its carboxyl third where they represent 42.5% of the sequence. The sequence PPPGMRPP is present three times in this region. This final third of the protein is also recognized by human sera containing antibodies to Sm. A fusion protein, encoded by λrb91 (Figure 1), which contains only the last 80 amino acids of N was probed with an Sm serum that recognizes the snRNP proteins A, B/B′, N and D on immunoblots. When the bound antibodies were eluted and used to reprobe immunoblots of snRNP proteins they bound to A, B/B′ and N but not D (Figure 4a and data not shown). Most likely this means that these proteins share an epitope which is contained within the final 80 amino acids of N. Indeed, N does share some sequence similarity with A (Figure 3) and considerable similarity with B/B′ (Ohosone *et al.*, in press) in this region. It seems likely, therefore, that this shared epitope is defined by the similar sequences found in A and N. Evidence supporting this conclusion comes from studies using polyclonal antibodies raised in rabbits to probe immunoblots of snRNP proteins. Antibodies raised against a peptide containing residues 216−238 of N recognize the proteins A, N and B/B′ on immunoblots (Figure 4b and c) and will also immunoprecipitate the Sm class of U snRNPs from ^{32}P-labelled PC-12 cells (Figure 5). In A, 14 of the 23 amino acids in this peptide are conserved (Figure 3), supporting the argument that these shared residues define an Sm epitope.

We also examined the expression of N in human brain. Immunoblots of total protein from human cortex probed with anti-Sm serum show bands corresponding to B and N but not to B′ (Figure 6a). An identical result was obtained when immunoblots of snRNP proteins immunoprecipitated from human cortex were probed in the same way (Figure 6b). However, because B′ and N are difficult to distinguish by SDS−PAGE, it was considered possible that B′ was being masked by N in our immunoblots. To confirm that B′ could be detected if it was present in human cortex, a mixed sample of HeLa cells (which express B/B′) and snRNP proteins immunoprecipitated from human cortex (which expresses B and N) were probed as before and three bands, B/B′ and N, could be clearly seen (Figure 6c).

Based on Northern and immunoblot data, we previously showed that N was distinct from B/B′ because mRNA species encoding N are only revealed in tissues that express N protein (McAllister *et al.*, 1988). As in rats, human N mRNA is 1.6 kb in length and is most abundant in brain and absent or expressed at low levels in liver. N mRNA is also present in some cell lines (PC-12, MTC, TT and Raji) but not others (A20-2J and HeLa) (Figure 7). However, when Northern blots of three human cell lines (TT, HeLa and Raji cells) were hybridized with a ^{32}P-labelled cDNA encoding N and washed at lower stringency two mRNA species were revealed. One was the expected 1.6 kb N mRNA present predominantly in TT cells, to a lesser extent

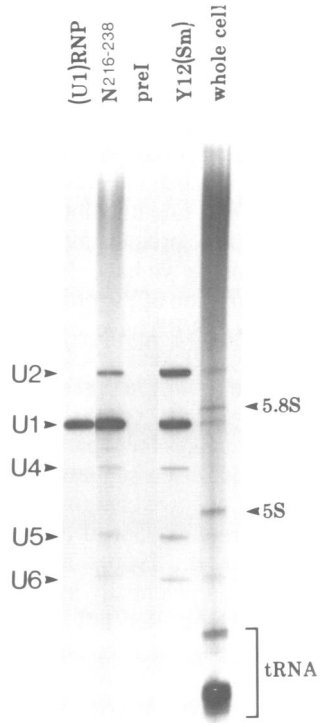

Fig. 5. RNA analysis of immunoprecipitates from ^{32}P-labelled PC-12 extracts. snRNPs were immunoprecipitated from cells using Sm monoclonal antibodies [lane Y12(Sm)], U1 specific human serum [lane (U1)RNP], rabbit polyclonal antibodies raised against an N peptide (lane N216−238) and rabbit pre-immune serum (lane preI). RNA was extracted from these samples and from the cell extract (lane whole cell) and analysed on a 10% acrylamide TBE−urea gel.

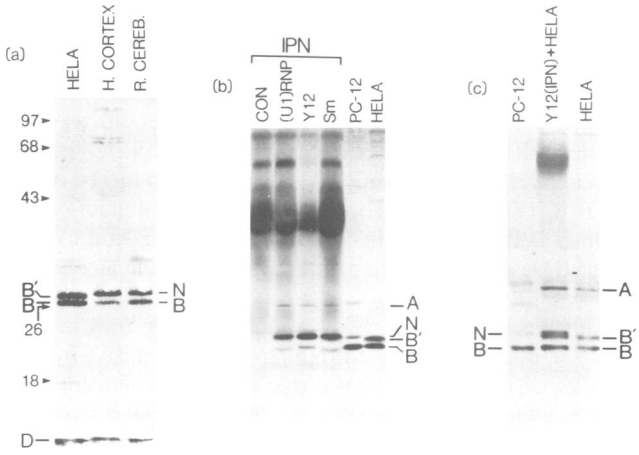

Fig. 6. Immunoblots of N in human brain. (**a**) Total protein from HeLa cells (lane HELA), human cortex (lane H. CORTEX) and rat cerebellum (lane R. CEREB.) probed with an anti-Sm serum. (**b**) Proteins immunoprecipitated from human cortex by non-autoimmune serum (lane CON), anti-(U1)RNP serum [lane (U1)RNP], mouse monoclonal anti-Sm antibodies (lane Y12) and anti-Sm serum (Sm) all probed with an anti-Sm serum. Total protein from PC-12 cells (lane PC-12) and HeLa cells (lane HELA) are also shown. (**c**) Total protein from PC-12 cells (lane PC-12), a mixture of Y12 immunopurified snRNPs from human cortex and total protein from HeLa cells [lane Y12(IPN)+HELA] and total protein from HeLa cells (HELA) probed with an anti-Sm serum to demonstrate that N and B′ can be distinguished on these blots. Pre-stained mol. wt markers (BRL) are shown on the left and the snRNP proteins A, N, B, B′ and D are indicated.

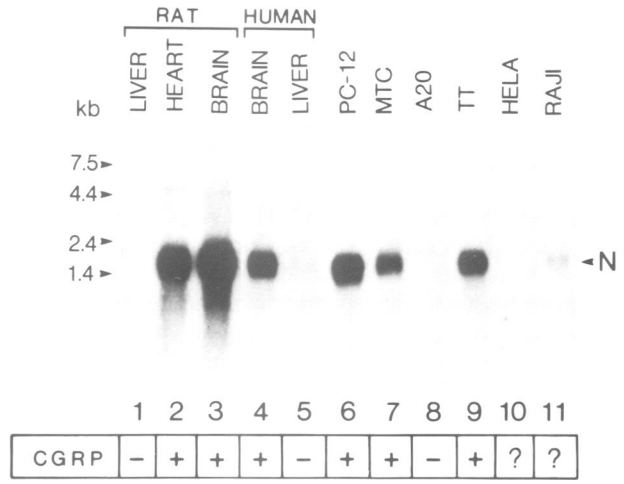

Fig. 7. Tissue and cell distribution of N mRNA Northern blots containing 1 μg of poly(A)$^+$ RNA from rat liver (**lane 1**), rat heart (**lane 2**), rat brain (**lane 3**), human brain (**lane 4**), human liver (**lane 5**), PC-12 cells (**lane 6**), MTC cells (**lane 7**), A20-2J cells (**lane 8**), TT cells (**lane 9**), HeLa cells (**lane 10**) and Raji cells (**lane 11**) were probed with a ^{32}P-oligolabelled rat N cDNA. The ability of these cells and tissues to splice the calcitonin/CGRP gene to CGRP mRNA is indicated by a (+) in the table below the Northern blot. These data are based on the work of Nelkin *et al.* (1984), Crenshaw *et al.* (1986), Leff *et al.* (1987) and Bennett and Amara (unpublished data).

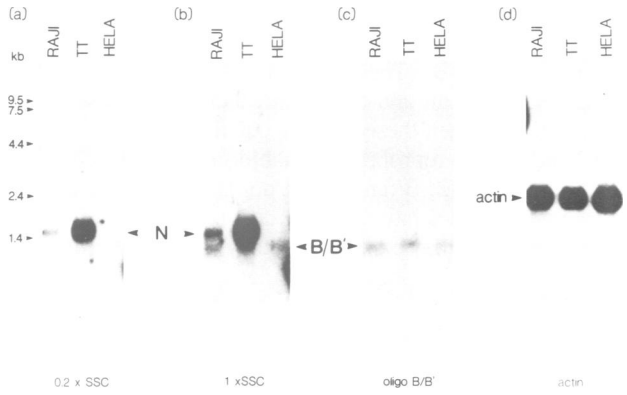

Fig. 8. Northern blot analysis of N. Northern blots containing 2 μg of poly(A)$^+$ RNA from Raji cells (lane RAJI), TT cells (lane TT) and HeLa cells were hybridized with various probes. The blots in (**a**) and (**b**) were probed with a ^{32}P-oligolabelled N cDNA and subsequently washed with 1 × SSC/0.1% SDS at 65°C (b) and then with 0.2 × SSC/0.1% SDS at 65°C (a). The same blot in (**c**) was then reprobed with a ^{32}P end-labelled oligonucleotide corresponding to B/B′ (Ohosone *et al.*, submitted). A parallel blot (**d**) was probed with a nick-translated actin cDNA to confirm that equivalent amounts of intact poly(A)$^+$ RNA were present in each lane. RNA size markers are shown to the left in kb and the mRNAs encoding N, B/B′ and actin are indicated.

in Raji cells, and not detected in HeLa cells while the other was a 1.3 kb mRNA present in all three cell lines (Figure 8b). As the stringency of washing was increased the lower 1.3 kb mRNA species was selectively lost from all three cell lines (Figure 8a). This lower mRNA species was confirmed to encode B/B′ because it hybridized to a B/B′ specific oligonucleotide (Figure 8c). A comparison of the cDNA sequences of N and B/B′ reveals that they are closely related but distinct (Schmauss *et al.*, in press). The most obvious difference between the two sequences is that B/B′ contains a 150 bp region in its coding sequence that is not found in N.

Therefore we used an oligonucleotide sequence from this region as a B/B′ specific probe. From immunoblot studies, TT cells were known to express N and all the cell lines express B/B′ (data not shown).

Discussion

We previously identified a rat snRNP-associated polypeptide designated N, which is expressed predominantly in brain. Figure 2 presents the nucleotide sequence of the cDNA encoding N. The deduced protein sequence of N is 240 amino acids long and has a predicted mol. wt of 24.614 kd. It is very proline-rich (>40% in the carboxyl-terminal third of the protein) and has no significant sequence similarity with D or E (Rokeach *et al.*, 1988; Wieben *et al.*, 1985), both of which are recognised by some Sm sera, or with 70K or B″ (Theissen *et al.*, 1986; Habets *et al.*, 1987).

A major reason for cloning snRNP proteins is that they are often recognized by autoimmune sera. Identifying the exact nature of the Sm epitopes characteristic of SLE for example, may provide some insight into how autoimmune diseases occur. We know that an epitope recognized by Sm sera is contained in the last 80 amino acids of N, and is expressed in bacteria, because anti-Sm sera recognize the fusion protein from clone λrb91 containing this region of N. The fact that this epitope is made in bacteria suggests that it is part of the primary sequence of N and not a post-translational modification. We also know that N shares an epitope recognized by anti-Sm sera with the snRNP proteins A and B/B′. The fusion protein encoded by λrb91 was probed with an anti-Sm serum that recognizes A, B, B′, N and D. When these affinity purified antibodies were eluted and used to reprobe immunoblots of snRNP proteins, they bound to A, B, B′ and N but not D (Figure 4 and data not shown). N does have some sequence similarity with A (Figure 3) which may explain why they share an epitope. N also shares extensive sequence similarity in this region with B/B′ (Ohosone *et al.*, in press). Antibodies raised against a peptide (residues 216−238 of N) containing part of this shared sequence (see Figure 3) recognize A, N and B/B′ but not D (Figure 4). It has previously been observed that although anti-A containing anti-RNP sera selectively precipitate U1 snRNP, they also react with B/B′ on immunoblots (e.g. Pettersson *et al.*, 1984). Therefore it was possible that the antibodies raised against this peptide recognized an RNP rather than an Sm epitope. However, this turned out not to be the case as these antibodies immunoprecipitate the Sm class of U snRNPs from ^{32}P-labelled PC-12 cells (Figure 5) suggesting that amino acids 216−238 of N define an Sm epitope. A similar immunoprecipitation pattern is observed when A20-2J cells, which express little or no N, are used (data not shown). Although all of the Sm snRNPs are immunoprecipitated in these experiments, the relative proportion of U1 snRNA appears to be greater with the rabbit serum compared to the Sm monoclonal Y12 [Figure 5; lanes N216−238 and Y12 (Sm)]. This suggests that the rabbit serum also recognizes A in the native U1 snRNP. The sequence PPPGMRPP is present twice in this peptide, and three times in all of N. This sequence is also present twice in B/B′ (Ohosone *et al.*, in press) and a very similar sequence is found in A, PPPGMIPP (Sillekens *et al.*, 1987). Interestingly, the snRNP protein C also shares some sequence similarity to N, mainly because it has a proline-rich region (Yamamoto *et al.*, 1988) and it also has a version of this sequence, PAPGMRPP. However, C is not recognized by anti-Sm containing serum, nor is it recognized by the anti-peptide antibodies raised in rabbits (Figure 4). There must be more than one Sm epitope however, because Sm sera usually recognize D and the sequence we describe as a potential Sm epitope is not present in D. Also, the mouse monoclonal anti-Sm antibody, Y12, recognizes B/B′ and D on immunoblots (Pettersson *et al.*, 1984) confirming that these proteins must share an epitope.

The potential Sm epitope we have defined in N is found within the proline-rich region of the protein. Several known autoantigens (70K, A, C, B/B′) have such proline-rich areas (Theissen *et al.*, 1986; Sillekens *et al.*, 1987; Yamamoto *et al.*, 1988; Ohosone *et al.*, in press) and perhaps the structures of these regions, or their possible role as RNA-binding domains makes them particularly likely to be autoantigenic. It has also been suggested that proline-rich sequences with overall positive charges may be features of nucleic acid binding proteins (Garoff *et al.*, 1980; Kruijer *et al.*, 1981; Adam *et al.*, 1986). This is an attractive conclusion for N as it can potentially interact with snRNAs as well as with pre-mRNAs. Recently, a 'RNP' consensus sequence representing part of a conserved RNA binding domain has been described (Swanson *et al.*, 1987) and it is present in the 70K and B″ snRNP proteins (Theissen *et al.*, 1986; Habets *et al.*, 1987). This sequence is not present in N, B/B′ (Ohosone *et al.*, in preparation), D (Rokeach *et al.*, 1988) or E (Wieben *et al.*, 1985). However, N may still bind RNA because the 'RNP' consensus is not an obligate feature of all RNA-binding proteins (Swanson *et al.*, 1987).

How similar is N to B/B′? Previous RNA blotting analyses showed a single 1.6 kb mRNA species only in those tissues or cell lines which express N (Figure 6; McAllister *et al.*, 1988). Therefore, we concluded that N is distinct from B/B′ which are thought to be expressed in all cells. Nonetheless, N and B/B′ share Sm epitopes (Figure 4), are of similar size and are snRNP proteins. Also, because the RNA blots were hybridized and washed at high stringency, it was possible that less stringent hybridization conditions would reveal a related mRNA species. Lower stringency hybridizations did reveal a related 1.3 kb mRNA species, which turned out to be B/B′ (Figure 8). A comparison of rat N, human N and human B/B′ confirms that N and B/B′ are indeed closely related but encoded by two different genes (Schmauss *et al.*, in press).

snRNPs are also studied because of their role in pre-mRNA processing. An intriguing question raised by our results is 'do snRNPs play a role in the mechanism(s) by which alternative splice sites are selected from a common pre-mRNA precursor, especially when this is carried out in a tissue-specific or developmentally regulated manner?' Two lines of evidence provide circumstantial support for such speculation. First, N is a tissue-specifically expressed snRNP protein. Also, immunoblots of human cortex probed with anti-Sm serum demonstrate that B′ is not expressed in this tissue, whereas B and N are (Figure 6). The similarity of sequence between N and B/B′ suggests that they may have similar functions within snRNP particles, and perhaps a change in the complement of snRNP proteins in different tissues is involved in tissue-specific splice choices. Second, genomic Southern blots demonstrate similar sequences to N in rat, human, mouse, cow, chicken and frog DNA (data not shown) indicating that N is expressed in vertebrates in

general. In humans, N is expressed in the same tissue-specific manner as it is in rats (Figures 6 and 7). That is, N is predominantly expressed in brain. Previously, we noted the interesting correlation between the expression of N in various cell lines and tissues and their ability to splice pre-mRNA transcribed from the calcitonin/calcitonin gene related peptide (CGRP) to CGRP mRNA (McAllister *et al.*, 1988). This correlation has now been extended to include human brain and TT cells as well as several rat cell lines and tissues (Figure 7).

In conclusion, we have demonstrated that in certain tissues the complement of proteins associated with snRNPs varies. It is tempting to speculate that these variations allow a change in the specificity of splice site recognition by the snRNPs. More specifically, in brain B′ may be replaced in snRNPs by N, a related protein, which can still carry out the basic functions of snRNPs in splicing but which could allow a change in splice site selection. This change could occur because the tissue-specific snRNPs carry out some splice choices less efficiently and therefore utilize others, or because they carry out new splice choices preferentially. The existence of cDNA clones for N and cell lines which can differentially splice the calcitonin/CGRP gene should allow us to test this hypothesis in a defined experimental system.

Materials and methods

Sequencing

DNA sequencing was performed using the chain termination method (Sanger *et al.*, 1977). Overlapping restriction fragments were subcloned into the M13 bacteriophage derivatives mp18 and mp19. All sequence analysis and comparisons were carried out using the software provided with the Genemaster DNA workstation (Biorad, Richmond, CA).

Cell lines

The following cell lines were used in these experiments: PC-12 (rat pheochromocytoma), MCT (6-23, rat medullary thyroid carcinoma), A20-2J (mouse B cell lymphoma), TT (human medullary thyroid carcinoma), HeLa (human cervical carcinoma) and Raji (human B cell lymphoma). All were obtained from the American Tissue Culture Collection, Rockville except TT cells which were kindly provided by Dr Susan Leong.

Immunoblotting and immunoprecipitations

Protein extraction and immunoblotting were carried out as described previously (McAllister *et al.*, 1988). Immunoprecipitation of snRNPs from tissues and cell lines was done essentially as described by Kessler (1975).

Northern hybridizations

Total RNA was extracted by the method of Chirgwin *et al.* (1979). Poly(A)$^+$ RNA was prepared by oligo(dT)–cellulose chromatography (Aviv and Leder, 1972), separated on 1.2% agarose gels containing formaldehyde (Lehrach *et al.*, 1977), and transferred to Zetaprobe membranes as recommended by the manufacturer (Biorad). N cDNA probes were made from a *Hin*dIII restriction fragment of clone pGMA2 containing the entire coding region and oligolabelled as described by Feinberg and Vogelstein (1983). A B/B′ specific oligonucleotide probe (CGGAGCTAC-TTCCATACTCTGTGG) was labelled using polynucleotide kinase (New England Biolabs), as described in Davis *et al.* (1986) and used to probe Zetaprobe membranes using the conditions recommended by the manufacturer (Biorad, Richmond). The γ-actin cDNA probe, provided by Dr L.Kedes was labelled by nick translation (Rigby *et al.*, 1977). Northern blots were prehybridized and hybridized in 5 × SSC, 1% SDS, 0.5% Blotto, 1 × Denhardt's, 20 mM sodium phosphate, pH 6.8, 100 μg salmon sperm DNA at 65°C.

Peptides and oligonucleotides

Peptides and oligonucleotides were synthesized at the Protein and Nucleic Acid Chemistry Facility, Yale University School of Medicine.

Raising antibodies

The peptide N (216–238) was coupled to keyhole limpet haemocyanin (KLH) in the presence of glutaraldehyde (Reichlin, 1980). Approximately 1 mg of the coupled peptide in incomplete Freund's emulsion was used to immunize a rabbit five times at 21-day intervals. Serum was collected 10 days after the final boost.

Acknowledgements

We are grateful to Joan Steitz and John Hardin for providing SLE sera and to Volker Gerke for the gift of semi-purified U1 snRNPs. We would like to thank Yasuo Ohosone and co-workers for communicating their results prior to publication and John Alsobrook for assistance with computers. We would also like to thank Randy Blakely, Jennifer Pinkham and Claudia Schmauss for critical reading of the manuscript.

References

Adam,S.A., Nakagawa,T., Swanson,M., Woodruff,T.K. and Dreyfuss,G. (1986) *Mol. Cell. Biol.*, **6**, 2932–2943.
Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408–1412.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Crenshaw,E.B., Russo,A.F., Swanson,L.W. and Rosenfeld,M.G. (1987) *Cell*, **49**, 389–398.
Davis,L.G., Dibner,M.D. and Battey,J.F. (1986) *Basic Methods in Molecular Biology*. Elsevier, New York.
Feinberg,A. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Garoff,H., Frischauf,A.M., Simons,K., Lehrach,H. and Delius,H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6376–6380.
Habets,W.J., Sillekens,P.T.G., Hoet,M.H., Schalken,J.A., Roebroek, A.J.M., Leunissen,J.A.M., van den Ven,W.J.M. and van Venrooij,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2421–2425.
Kessler,S.W. (1975) *J. Immunol.*, **115**, 1619–1624.
Kozak,M. (1986) *Cell*, **42**, 283–292.
Kruijer,W., van Schaik,F.M.A. and Sussenbach,J.S. (1981) *Nucleic Acids Res.*, **9**, 4439–4457.
Leff,S.E., Evans,R.M. and Rosenfeld,M.G. (1987) *Cell*, **48**, 517–524.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4751.
Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495–5499.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Luhrmann,L. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor snRNPs*. Springer, Heidelberg, pp. 71–99.
McAllister,G., Amara,S.G. and Lerner,M.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5296–5300.
Nelkin,B.D., Rosenfeld,K.I., de Bustros,A., Leong,S.S., Roos,B.A. and Baylin,S.B. (1984) *Biochem. Biophys. Res. Commun.*, **123**, 648–655.
Ohosone,Y., Mimori,T., Griffith,A., Akizuki,M., Homma,M., Craft,J. and Hardin,J.A. (1989) *Proc. Natl. Acad. Sci. USA*, in press.
Pettersson,I., Hinterberger,M., Mimori,T., Gottlieb,E. and Steitz,J.A. (1984) *J. Biol. Chem.*, **259**, 5907–5914.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Reichlin,M. (1980) *Methods Enzymol.*, **70**, 159–165.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Rokeach,L.A., Haselby,J.A. and Hoch,S.O. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4832–4836.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schmauss,C., McAllister,G., Ohosone,Y., Hardin,J.A. and Lerner,M.R. (1989) *Nucleic Acids Res.*, in press.
Sillekens,P.T.G., Habets,W.J., Beijer,R.P. and van Venrooij,W.J. (1987) *EMBO J.*, **6**, 3841–3848.
Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor snRNPs*. Springer, Heidelberg, pp. 115–154.
Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.*, **7**, 1731–1739.
Tan,E.M. (1982) *Adv. Immunol.*, **33**, 167–240.
Theissen,H., Etzerodt,M., Reuter,R., Schneider,C., Lottspeich,F., Argos,P., Luhrmann,R. and Philipson,L. (1986) *EMBO J.*, **5**, 3209–3217.
Wieben,E.D., Rohleder,A.M., Nenninger,J.M. and Pederson,T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7914–7918.
Yamamoto,K., Miura,H., Moroi,Y., Yoshinoya,S., Goto,M., Nishioka,K. and Miyamoto,T. (1988) *J. Immunol.*, **140**, 311–317.

Received on September 30, 1988; revised on January 9, 1989