Volume 17 Number 4 1989 Nucleic Acids Research

A comparison of snRNP-associated Sm-autoantigens: human N, rat N and human B/B'

Claudia Schmauss*, George McAllister, Yasuo Ohosone[1], John A.Hardin[1] and Michael R.Lerner

Section of Molecular Neurobiology, Howard Hughes Medical Institute Research Laboratories and [1]Department of Internal Medicine, Section of Rheumatology, Yale University School of Medicine, 333 Cedar Street, PO Box 3333, New Haven, CT 06510, USA

Received September 22, 1988; Revised and Accepted January 16, 1989

ABSTRACT

N is a tissue-specific, Sm-epitope bearing, snRNP-associated protein found predominantly in brain. The cDNA sequence encoding human N is compared to those for rat N and human B/B'. The amino acid sequences of human and rat N are 100% conserved. Although the amino acid sequences of N and B/B' are very similar to each other, B/B' contains 50 amino acids which are not present in N. On Northern blots the cDNAs encoding N and B/B' recognize two different RNA species. A comparison of the codon usage, as specified by the open reading frames of N and B/B' as well as results from Southern blots, show that N and B/B' are derived from different genes.

INTRODUCTION

The major small nuclear ribonucleoproteins (snRNPs) are particles composed of discrete sets of proteins associated with the small nuclear RNAs U1, U2, U5, and U4/U6. Each human snRNP is generally thought to contain at least seven common polypeptides called B, B', D, D', E, F and G (1). Since the difference between B and B' is unclear and since the two proteins may even be encoded by mRNAs derived from the same gene (2) they will be referred to here as B/B'. Some snRNPs also contain characteristic proteins such as the U1 snRNP-associated polypeptides 70K, A, and C, and the U2 snRNP-associated polypeptides A' and B" (1).

The general roles of snRNPs in pre-mRNA splicing are known. For example, the U1 snRNP acts at the 5' splice site, the U2 snRNP interacts with the branch point, and the U5 snRNP probably associates with the 3' splice site (for review see 3). However, the specific functions of the individual snRNP polypeptides in this process remain to be elucidated. Some of these proteins are also important because patients with the autoimmune disease systemic lupus erythematosus (SLE) have autoantibodies directed against them (4, 5). The most common autoantigen is

called Sm and the presence of circulating autoantibodies to it is pathognomonic for SLE. Such autoantibodies can precipitate the U1, U2, U4, U5, and U6 snRNPs from cell extracts because they all contain B/B' which bear Sm epitopes (2).

Recently, a snRNP associated polypeptide called N, was identified in rats (6). N is recognized by anti-Sm sera and is of similar molecular weight to B/B' as estimated by sodium dodeccyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). However, unlike other snRNP proteins, it is present only in certain tissues. It is most abundant in brain, less so in heart and undetectable in liver. N may be involved in tissue-specific splicing events (6).

N and B/B' are encoded by two distinct but related mRNA species. However, sequence comparisons of N and B/B' have been complicated because N was characterized from rat tissue (McAllister et al., submitted) while B/B' was characterized from a human cell line (7). It is unclear whether their mRNAs are the products of alternative splicing of pre-mRNA transcribed from the same gene or are from two distinct but related genes. To make accurate comparisons between N and B/B', they must be compared in the same species.

MATERIALS AND METHODS

1. Library screening for cDNA clones encoding N

A lambda gt11 human cerebellar cDNA library (8, obtained from American Type Culture Collection) was screened with a ^{32}P-radiolabeled Hind III fragment of brain cDNA which contains the entire coding region of rat N (McAllister et al., submitted). Positive clones were purified by standard methods (9).

2. DNA sequencing

DNA sequence analysis was performed by the dideoxy chain termination method of Sanger et al (1977) (10) in conjunction with the single stranded M13 vectors mp19 and mp18. The sequencing strategy used to determine the complete cDNA sequence of the human cerebellar clone hcVI1 is illustrated in Fig.1. Unless otherwise indicated the M13 mp19 vector was used. The DNA sequence and its deduced amino acid sequence was compiled using the Gene-Master DNA Workstation software (Bio-Rad Laboratories). The nucleotide sequence of human cerebellar N and its deduced amino acid sequence were compared with those in the

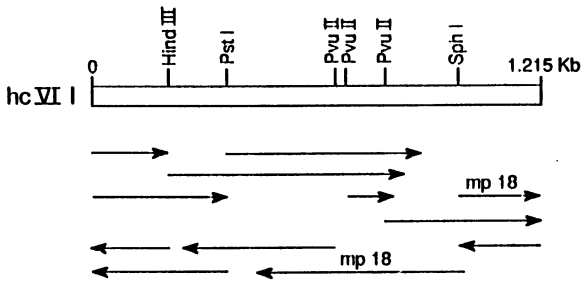


Figure 1.
Restriction map and sequencing strategy of hcVI1 cDNA.
The DNA sequence was determined by the dideoxy chain terminator method (9) using the single-stranded M13mp19 vector unless otherwise indicated. The complete nucleotide sequence of the hcVI1 Eco RI restriction fragment was obtained in both orientations from overlapping restriction fragments as indicated by the arrows. Arrowheads show the direction of sequencing.

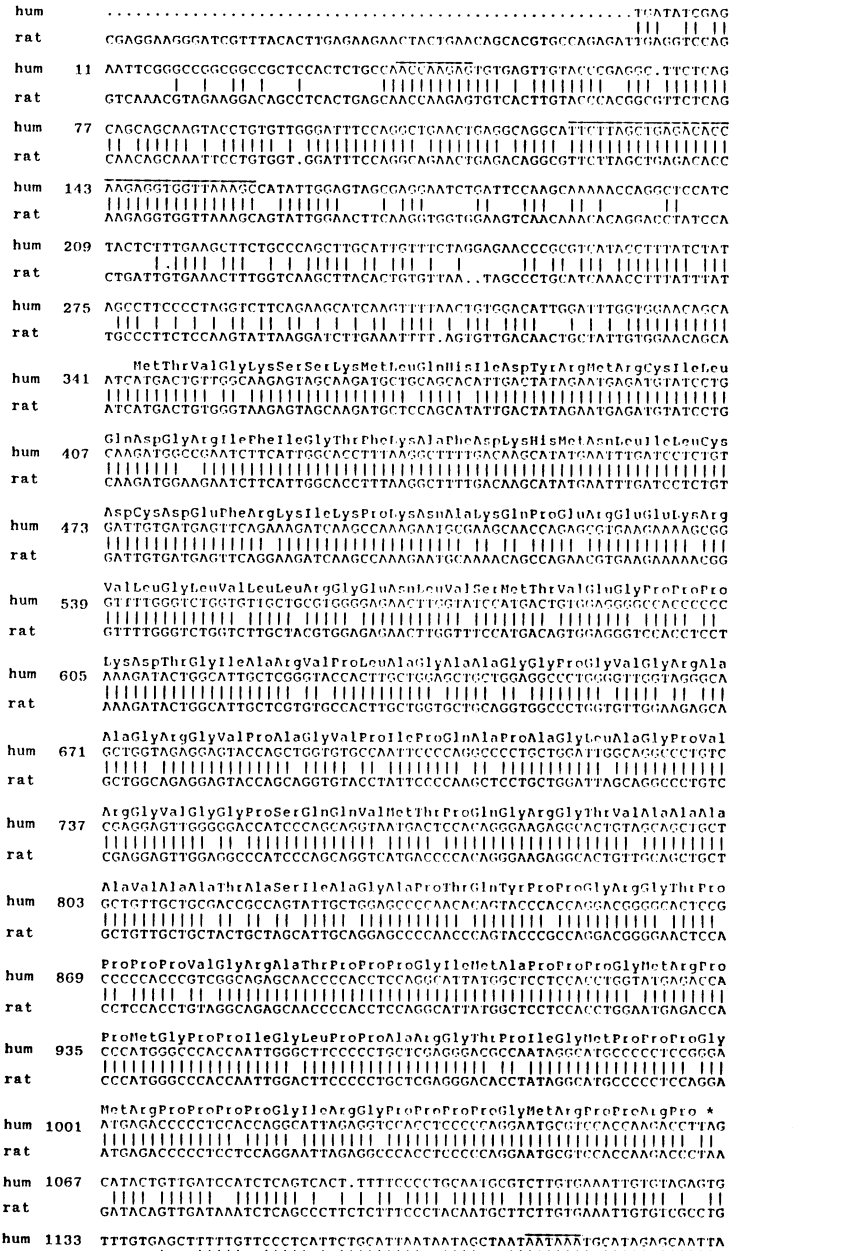
Genbank database. All nucleotide and amino acid sequence comparisons were performed using the alignment algorithm of Needleman and Wunsch (Sequence Analysis Software Package, University of Wisconsin and GeneMaster DNA Workstation software, Bio-Rad Laboratories).

3. Northern blot analysis

Analyses of human N and B/B' mRNA species were carried out by size fractionation of oligo dT cellulose selected RNA in formaldehyde-agarose gels, followed by transfer to nylon membranes (11). The membranes were prehybridized in a solution containing 5x SSPE, pH 7, 5x Denhardt's solution, 0.1M sodium phosphate, pH 6.8, 0.1% SDS, 0.005M EDTA, 20 ng/ml Poly (A)n, 100 ug/ml salmon sperm DNA, and 100 ug/ml yeast RNA. The total length cDNAs encoding N and B/B' were labelled with ^{32}P by random priming (12) and hybridized at 65°C overnight to the immobilized RNA in a buffer of 5x SSC, pH 7, 1x Denhardt's solution, 0.02 M sodium phosphate, pH 6.8, 0.2% SDS, 0.005M EDTA, and 20 ug/ml Poly (A)n. Membranes were washed in 0.2x SSPE at 65 ° C.

4. Southern blot analysis

For Southern blot analysis, 10 ug of human genomic DNA was digested with the restriction enzymes EcoRI and Hind III, separated by electrophoresis in 1.2% agarose gels, and transferred to a Zeta-Probe membrane



(Bio-Rad). Hybridization of specific DNA fragments to the human cDNAs encoding N and B/B' was conducted at 65°C with 5x SSPE, 5x Denhardt's Solution, 1% SDS, 100 ug/ml salmon sperm DNA and 100 ug/ml yeast RNA.

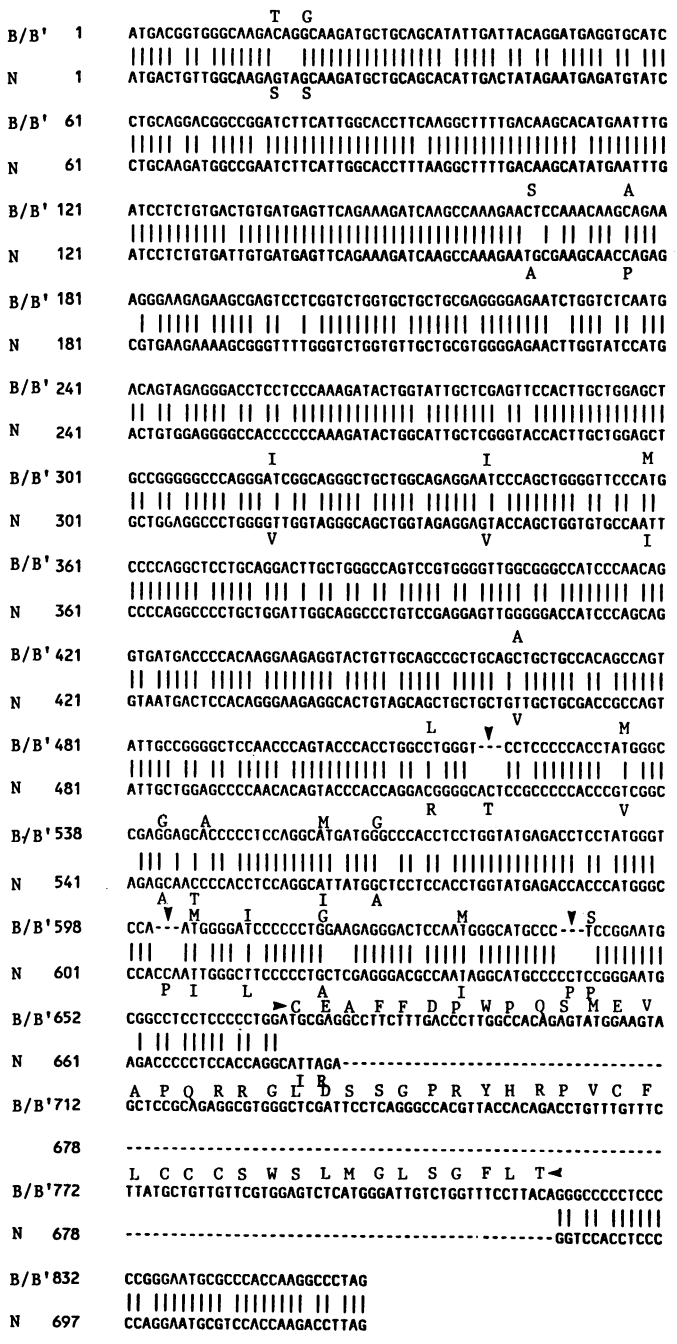
RESULTS AND DISCUSSION

50,000 recombinant lambda gt11 phage from a cDNA library containing inserts representing mRNAs extracted from human cerebellum were probed with a ^{32}P-radiolabeled HindIII restriction fragment containing the entire coding region of rat brain N. Of ten positive clones, one, called hcVI1, was subcloned into the vector Bluescript (Stratagene) and further characterized by restriction mapping. It was found to contain an insert of approximately 1,200 nucleotides, the restriction map of which was similar to that obtained for the rat N cDNA (McAllister et al., submitted).

The 1,215 base nucleotide sequence of the cDNA insert of clone hcVI1 and its deduced amino acid sequence are shown in Fig. 2. There is a single large uninterrupted open reading frame (ORF) of 720 nucleotides encoding 240 amino acids. The first ATG codon occurs at position 344 and is designated as the translation initiation codon because it exists within the sequence ATCATGA which diverges from the consensus sequence for initiation (13) only at positions -3 and +4 (see Fig. 2). A TAG translation stop codon is present at position 1,064 and is followed by a 3' nontranslated region containing the polyadenylation signal AATAAA which is in turn followed by a Poly(A) tail.

A Needleman-Wunsch nucleotide alignment of rat and human cDNA sequences for N revealed an overall homology of 86% while the nucleotides for the amino acid coding region are 91.2% homologous. Here, 60 of 720 nucleotides diverge and 59 of these differences occur at the third base position in codons as specified by the ORF. The predicted amino acid sequence of rat and human N are identical (Fig.2).

Figure 2.
Comparison of the complete nucleotide sequence of hcVI1 with the nucleotide sequence of N obtained from a rat cDNA clone (McAllister et al., submitted). Alignment was achieved by the algorithm of Needleman and Wunsch and as few deletions or insertions as possible are assumed. The predicted amino acid sequence of N is shown above the nucleotide triplets of hcVI1 and is identical in humans and rats.



Whereas the 5' and 3' noncoding regions are not as precisely conserved as the coding region, spans of homologous nucleotides are frequent. The most distinctive one is a run of 33 conserved nucleotides in the 5' untranslated region: TTCTTAGCTGAGACACCAAGAGGTGGTTAAAGC. Part of this conserved sequence -ACCAAGAG- also appears as a repeat 91 nucleotides further upstream (see Fig.2).

The complete conservation of N between the two species is striking, but considering that snRNPs are involved in the evolutionarily old process of pre-mRNA splicing, all snRNP proteins are probably highly conserved. These proteins probably have multiple sites for protein/protein and protein/RNA interactions which hold the particles together and enable them to form components of the splicing machinery (14). It may be that these interactions impose constraints on the likelihood of snRNP proteins to diverge during the course of evolution causing them to be more highly conserved than other ubiquitous proteins which are not as interconnected with other macromolecules. Histones, which have multiple sites of interactions with each other and their surrounding nucleic acid, may represent an analogous situation (15-17). For example, H4 of peas and cows differ in only 2 out of 102 amino acid residues (16). For N, the potential role in tissue-specific pre-mRNA processing events might lead to additional evolutionary pressure to retain its sequence from one species to the next.

How related are the human Sm-autoantigens N and B/B'? A comparison of the human amino acid sequence for N with that of B/B' also reveals remarkable similarities. The first 226 amino acids of each clone are 90.3% identical and 18 of the 19 pairs of divergent amino acids represent conservative changes. However, three amino acids, two prolines and a threonine present in N (at positions indicated in Fig.3 by arrowheads) are missing in the aligned positions of B/B'. Two further amino

Figure 3.
Comparison of the predicted amino acid sequences of human N and B/B' based on the nucleotide sequences of their coding regions by Needlemann-Wunsch alignment. For their amino acid sequences, only the differences between N and B/B' are presented. Arrowheads indicate the most significant differences between the two sequences. The first 226 amino acids of each clone are 90.3% identical, while the twelve amino acids at the carboxy-terminus are common to both N and B/B'. The amino acid sequence of B/B' also contains additional 50 amino acids located towards the carboxy-terminus of the protein which are absent from N.

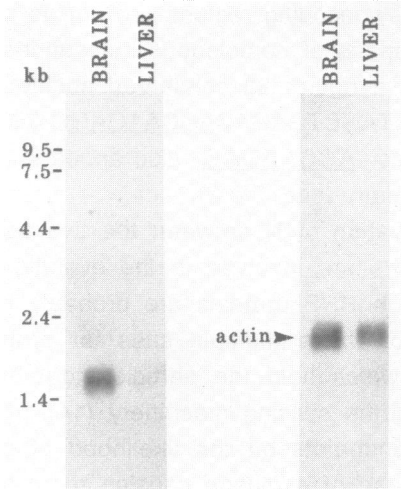


Figure 4.
Northern blot analysis of poly(A)-RNA from human brain and liver was carried out as described in Materials and Methods. One ug of poly(A)-RNA was loaded in each lane and the full length cDNA encoding N was used as a probe. For an internal control the membrane was reprobed with a 1.2 kb cDNA encoding actin. The size markers are derived from an RNA ladder (BRL). The autoradiographs were developed after a 12 hour exposure.

acids, isoleucine and arginine, found in the human N sequence, are missing in the aligned position of B/B', and mark the most substantial difference between N and B/B'. Here, B/B' contains an additional 50 amino acid insert not found in N (Fig.3, indicated by arrowheads). Nonetheless, the carboxy-terminal amino acid sequence GPPPPGMRPPRP, is common to both proteins (Fig.3). Interestingly, the U1-specific snRNP-associated protein A (18) and the U2-specific snRNP-associated protein B" (19) have a similar pattern of homology compared with that between N and B/B'. They have their highest homology in their carboxy-parts (86%) and a somewhat less extensive homology at their amino-parts(77%) (18). These homologous regions are separated by 103 and 40 amino acids for A and B", respectively.

On Northern blots the cDNA encoding human N hybridizes to a single 1.6 kb long RNA species prominent in human brain and which is either not present or present at a barely detectable level in human liver (Fig.4). mRNA for N is also found in the human cell lines Raji (Burkitt lymphoma) and TT (medullary thyroid carcinoma) but not in HeLa (cervical carcino-

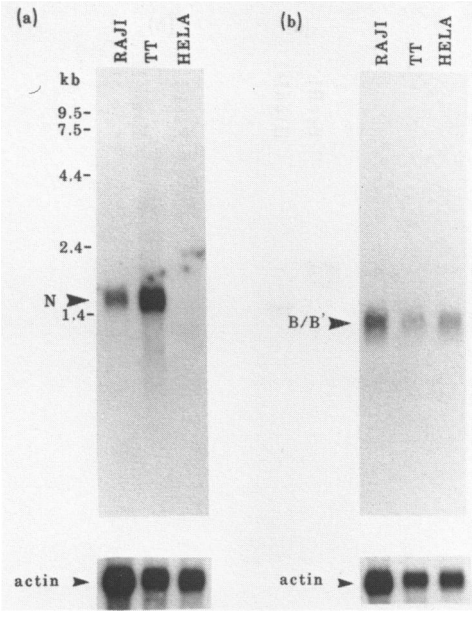


Figure 5
Northern blot hybridization of poly(A)-RNA from human cell lines. Two ug of RNA were loaded on each lane. (a) shows the membrane blot probed with full length cDNA encoding human N, (b) shows a membrane blot probed with full length cDNA encoding B/B'. Membranes were reprobed with an actin cDNA probe (see Fig.3). Size markers are derived from an RNA ladder (BRL). Membranes were exposed to the film for 2 days.

ma) (Fig.5). N mRNA is much more abundant in TT cells than in Raji cells. These results are identical to those obtained when the same tissues and cell lines are probed with the rat N cDNA (McAllister et al., submitted). On Northern blots the cDNA encoding B/B' recognizes an RNA species of 1.3 kb found in all three human cell lines Raji, TT and HeLa, when the membrane is hybridized and washed at the same high stringency as the Northern blot probed with cDNA encoding N (Fig.5). It has been shown that at a lower stringency N also recognizes this RNA species (McAllister et al., submitted).

It is unclear why B/B' migrates faster than N on SDS-PAGE, as its deduced amino acid sequence is longer than that of N. Perhaps the presence of the 50 amino acids in B/B' which are absent in N leads to distinct structural differences or, post-translational modifications that produce

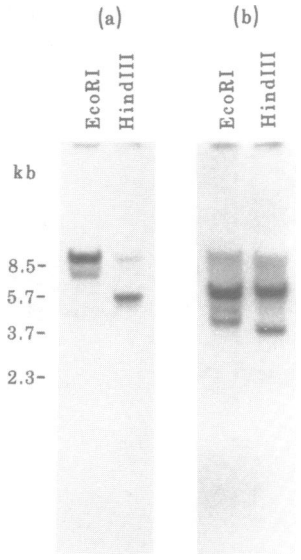


Figure 6.
Southern blot analysis of N and B/B' sequences in the human genome. Total human genomic DNA (10ug) was digested with the restriction enzymes EcoRI or HindIII. The DNA was transferred to a Zeta-Probe blotting membrane and a ^{32}P-radiolabeled cDNA encoding N (panel a) or a ^{32}P-labelled cDNA encoding B/B' (panel b) was hybridized to it as described in Materials and Methods. The size markers are from BstE II-digested bacteriophage lambda DNA. Numbers alongside the autoradiograph indicate the size of the fragments in kilobase pairs. The autoradiograph was developed after a 2 day exposure.

the unexpected migrations. We are presently working to clarify this issue.

In summary, N and B/B' recognize two distinct RNA species on Northern blots. In addition to occasional amino acid differences - additions, or deletions, throughout the otherwise homologous amino 2/3 of the coding regions for N and B/B'- the sequence for B/B' contains an insert of 50 amino acids near its carboxy-terminus that is not found in N. Two findings indicate that N and B/B' are derived from different genes. First, analysis of the cDNA sequences coding for human N and B/B' reveals different codon usages for identical amino acids throughout their coding regions (Fig.5). Second, full length cDNA probes encoding N and B/B' hybridize to different EcoRI and HindIII fragments of the human genome on

Southern blots (Fig.6). Thus, the differences between N and B/B' are not due to alternative splicing of one gene but rather to the presence of two related genes, at least one of which, N, gives rise to a protein that is perfectly preserved from rats to humans.

*To whom correspondence should be addressed

REFERENCES

1. Bringmann,P. & Luhrmann,R. (1986) Embo J. 5, 3509-3516
2. Reuter,R., Rothe,S. & Luhrmann,R. (1987) Nucl. Acids Res. 15, 4021-4034
3 Steitz,J.A., Black,D.L.,Gerke,V., Parker,K., Kraemer,A., Frendewey,D. & Keller,W. (1988) In: Birnstiel,M.L. [Ed.], Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Springer Verlag.
4. Lerner,M.R. & Steitz,J.A. (1979) Proc. Natl. Acad. Sci. USA 76, 5495-5499
5. Tan ,E.M. (1982) Adv. Immunol. 33, 167-240
6. McAllister, G., Amara, S.G. & Lerner,M.R. (1988) Proc. Natl. Acad. Sci. USA 85,5296-5300
7. Ohosone,Y., Mimori,T., Griffith,A., Akizuki,M., Homma,M.,Craft,J.& Hardin,J.A. (1989) Proc. Natl. Acad. Sci. USA, in press
8. Kamholz,J., de Ferra,F., Puckett,C. & Lazzarini, R. (1986) Proc. Natl. Acad. Sci. USA 83, 4962-4966
9. Maniatis,T., Fritsch,E.F., Sambrook,J. (1982) Molecular Cloning. Cold
10.Spring Harbor Laboratory.
Sanger, F., Nicklen,S., & Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
11.Church,G.M. & Gilbert,W. (1984) Proc. Natl. Acad. Sci. 81, 1991-1995
12.Feinberg,A., Vogelstein,B. (1984) Anal. Biochem. 137, 266-267
13.Kozak,M. (1986) Cell 44, 283-292
14.Hamm,J., Kazmaier,M. & Mattaj,I.W. (1987) Embo J. 6, 4379-4385
15.Grunstein,M., Schedl,P. & Kedes , L.M. (1976) J. Mol. Biol. 104, 351-369
16.Kedes, L.H. (1979) Ann. Rev. Biochem. 48, 837-870
17.Klug,A., Rhodes,D., Smith,J. , Finch, J.T. & Thomas, J.O. (1980) Nature 287, 509-516
18.Sillekens, P.T.G., Habets,W.J., Beijer,R.P. & van Venrooij, W.J. (1987) Embo J. 6, 3841-3848
19 Habets, W.J., Sillekens,P.T.G., Hoet,M.H., Schalken, J.A. Roebroek,A.J.M., Leunissen,J,A.M., van de Ven,W.J.M. & van Venrooij, W.J. (1987) Proc. Natl. Acad. Sci. USA 84, 2421-242