

## A chicken hnRNP of the A/B family recognizes the single-stranded d(CCCTAA)<sub>n</sub> telomeric repeated motif

Eleonora Marsich<sup>1</sup>, Antonella Bandiera<sup>1</sup>, Gianluca Tell<sup>1</sup>, Andrea Scaloni<sup>2</sup> and Giorgio Manzini<sup>1</sup>

<sup>1</sup>Department of Biochemistry, Biophysics, and Macromolecular Chemistry, University of Trieste, Italy;

<sup>2</sup>I.A.B.A.M., Centro Internazionale Servizi di Spettrometria di Massa, CNR, Napoli, Italy

With the aim of identifying proteins able to interact with the C-rich single-stranded telomeric repeated motif, three nuclear polypeptides, CBNP $\alpha$ , CBNP $\beta$  and CBNP $\gamma$ , with apparent mobilities in SDS/PAGE of 38, 44 and 55 kDa, respectively, were isolated from mature chicken erythrocytes by affinity chromatography. *In situ* UV-cross-linking experiments demonstrated that CBNP $\alpha$  and CBNP $\gamma$  interact directly with the telomeric d(CCCTAA)<sub>n</sub> repeat, whereas CBNP $\beta$  does not. Moreover, they provided information on the protein components responsible for each electrophoretic mobility-shift assay signal. Ion spray and matrix-assisted laser desorption ionization MS allowed us to identify CBNP $\alpha$  with single-stranded D-box-binding factor (ssDBF), a protein previously characterized as a transcription factor belonging to the A/B family of heterogeneous nuclear ribonucleoproteins, and CBNP $\beta$  with an isoform of this protein containing an extra exon. Similarly, CBNP $\gamma$  was shown to be probably the chicken homolog of hnRNP K, a ribonuclear protein able to bind to polyC oligonucleotides. The relation of CBNP $\alpha$  (i.e. ssDBF), CBNP $\beta$  and CBNP $\gamma$  to a number of similar proteins in the protein and nucleotide sequence databank is discussed. A rather diversified spectrum of functional roles has been assigned to some of these proteins despite the strong sequence homology among them.

**Keywords:** heterogeneous ribonucleoproteins (hnRNPs); nuclear proteins; ssDNA recognition; telomeric C-rich motif.

Since the discovery of the special features of repetitive DNA which constitutes the majority of telomeres of eukaryotic organisms, and of its specific replication machinery, considerable attention has also been directed to identifying and characterizing nuclear proteins that specifically bind to this DNA, in the normal Watson–Crick duplex form, as well as to the protruding single-stranded G-rich 3' overhang at the telomere terminus. Many proteins such as TRF1 and TRF2 that bind to duplex telomeric DNA in mammalian cells [1] and Rap1p in yeast [2] have been reported to be involved in telomere length maintenance and regulation of telomerase activity. In unicellular organisms, several proteins, such as  $\alpha\beta$  protein from *Oxytricha* [3], TBP from *Euplotes* [4], TEP and TGP from *Tetrahymena* [5,6], GBP from *Chlamydomonas* [7], which are known to interact with the single-stranded 3'-ending motif of the telomeres, have been characterized. Another group of proteins, ST-2 from *Trypanosoma* [8], qTBP42 from rat [9], human replication factor C [10], and murine STBP [11] and

A1/UP1 [12], also bind to the single-stranded G-rich telomeric motif, although their function has not been fully ascertained. The last of these, however, is the first ssDNA-binding protein shown to be directly involved in mammalian telomere biogenesis, suggesting a possible mechanism by which telomere length can be modulated [13]. Much less attention has been given to identifying nuclear components able to recognize the complementary single-stranded C-rich DNA repeat. Interestingly, a protein from *Trypanosoma*, ST-1, binds to the telomeric double-stranded repeat as well as to its single-stranded C-rich component [14]. In vertebrates, the nuclear protein from rat hepatocytes, qTBP42, has been shown to recognize each of the single-stranded forms of the telomeric repeat [9]. Several proteins that interact with polypyrimidine ssDNA have been described. They include NOGA4 in mouse [15] and rat [16] and the human hnRNP K protein, which binds to the single-stranded CT element of the *c-myc* gene promoter [17]. While this article was being completed, it came to our attention that hnRNP K and the splicing factor ASF/SF2 from HeLa nuclear extracts are able to recognize the single-stranded C-rich telomeric repeat [18]. These authors pointed out that other so far unidentified nuclear proteins could share this property. In this context it may be relevant to mention that earlier reports from this laboratory showed that proteins present in nuclear extracts from several vertebrate sources bind to the single-stranded telomeric repeat motif (CCCTAA)<sub>n</sub>, one of them exhibiting high sequence specificity [19,20]. This protein component does not recognize the complementary d(TTAGGG)<sub>n</sub> nor the telomeric duplex. We report here on the isolation and molecular identification of these nuclear proteins from extracts of mature chicken erythrocytes.

*Correspondence to* G. Manzini, Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy. Fax: + 39 040 6763691, Tel.: + 39 040 6763677, E-mail: manzini@bbcm.univ.trieste.it  
*Abbreviations:* CBNP, C-block-binding nuclear protein; hnRNP, heterogeneous nuclear ribonucleoprotein; EMSA, electrophoretic mobility shift assay; ISMS, ion-spray mass spectrometry; MALDIMS, matrix assisted laser desorption induced mass spectrometry; MSMS, tandem mass spectrometry; ssDBF, single-stranded D-box-binding factor.  
*Enzymes:* endoproteinase Asp-N (EC 3.4.24.33); endoproteinase Glu-C (EC 3.4.21.19).

(Received 14 July 2000; revised 26 October 2000; accepted 30 October 2000)



performed in batch: extracts ( $10\text{--}20\text{ mg}\cdot\text{mL}^{-1}$ ) diluted with 50 mM Tris/HCl, pH 8, containing 5 mM EDTA, 20% (v/v) glycerol, 0.5 mM dithiothreitol and 0.5 mM phenylmethanesulfonyl fluoride, to a final concentration of 200 mM NaCl. Generally, 40 mL diluted extract was incubated with 400  $\mu\text{L}$  affinity resin overnight at 4 °C on a rotor wheel. After incubation, the resin was washed several times with 50 mM Tris/HCl, pH 8, containing 5 mM EDTA and 200 mM NaCl, and the bound proteins were eluted with 10 mM Hepes, pH 7.9, containing 0.1 mM EGTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethanesulfonyl fluoride and 1 M NaCl for 1 h at room temperature.

### Reverse-phase chromatography

Proteins were fractionated by RP-HPLC on a Jupiter C<sub>18</sub> column (150 × 4.6 mm; 5  $\mu\text{m}$ ; 300 Å pore size; Phenomenex, Torrance, CA, USA) by using a linear gradient from 0% to 100% of acetonitrile in 0.1% trifluoroacetic acid over 40 min, at a flow rate of 0.5 mL·min<sup>-1</sup>. Individual components were collected manually and lyophilized.

### In-gel digestion and peptide purification

Proteins from SDS/PAGE were excised from the gel, triturated and washed with acetonitrile and 0.1 M NH<sub>4</sub>HCO<sub>3</sub> (twice each one). The proteins were in-gel reduced with 10 mM dithiothreitol in 0.1 M NH<sub>4</sub>HCO<sub>3</sub> (45 min, at 55 °C) and S-alkylated with 55 mM iodoacetamide in 0.1 M NH<sub>4</sub>HCO<sub>3</sub> (30 min at 25 °C and in the dark). After extensive washing with 0.1 M NH<sub>4</sub>HCO<sub>3</sub>, gel particles were dried and shrunk with a digestion solution of 12.5 ng· $\mu\text{L}^{-1}$  trypsin in 0.1 M NH<sub>4</sub>HCO<sub>3</sub>. After 2 h of incubation at 5 °C, an identical aliquot of digestion solution was added to the samples which were incubated overnight at 37 °C. Digestion solution was completely removed, and the gel pieces were sonicated with 50% acetonitrile in 0.1 M NH<sub>4</sub>HCO<sub>3</sub> (twice). The recovered solutions were mixed together and lyophilized. Peptide mixtures were fractionated by RP-HPLC on a Vydac C<sub>18</sub> column 214TP52 (250 × 2.1 mm; 5  $\mu\text{m}$ ; 300 Å pore size; The Separation Group, Hesperia, CA, USA) by using a linear gradient from 5% to 60% of acetonitrile in 0.1% trifluoroacetic acid over 65 min, at a flow rate of 0.2 mL·min<sup>-1</sup>. Individual components were collected manually and dried in a Speed-vac centrifuge (Savant). Similar experiments were performed by using endoproteases Asp-N and Glu-C. Digestions with Asp-N were carried out at 25 °C in 50 mM Tris/HCl, pH 7.5, for 16–20 h. Two independent digestions were set up with endoprotease Glu-C in 25 mM ammonium acetate, pH 4, and in 25 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 7.9, carrying out the reactions at 25 °C for 16–20 h.

### Mass spectrometry

Intact proteins were subjected to ion-spray MS (ISMS) analysis, using an API/SCIEX 100 ion-spray mass spectrometer. Samples were dissolved in acetonitrile/water (1 : 1) plus 0.1% (v/v) formic acid and then injected at a flow rate of 2  $\mu\text{L}\cdot\text{min}^{-1}$ . The quadrupole was scanned in the range  $m/z$  850–1200 and the spectra were acquired and elaborated using a manufacturer's program. Mass scale calibration was carried out using the multiple-charged ions of a separate introduction of myoglobin. All data are shown as average masses.

Matrix-assisted laser desorption ionization (MALDI) mass spectra were recorded by using a Voyager DE MALDI-TOF spectrometer (Perkin-Elmer-Perseptive Biosystem, Norwalk,

CT, USA); a mixture of analyte solution,  $\alpha$ -cyano-4-hydroxycinnamic acid and bovine insulin were applied to the sample plate and then dried. Mass calibration was performed using the molecular ions from bovine insulin (5734.59  $m/z$ ) and the  $\alpha$ -cyano-4-hydroxycinnamic acid (379.06  $m/z$ ) as internal standards. Raw data were analysed by using the OPUS software program provided by the manufacturer and are reported as average masses. MALDI MSMS spectra were recorded using an Autospec OA-TOF instrument (Micromass, Manchester, UK). In this case 2,5-dihydroxybenzoic acid was used as a matrix. The molecular ions collided with argon in a collision cell floated at 800 eV. The masses of fragment ions were measured using an orthogonally mounted TOF analyser. Data analysis was carried out by using the OPUS software supplied with the instrument.

### Protein identification

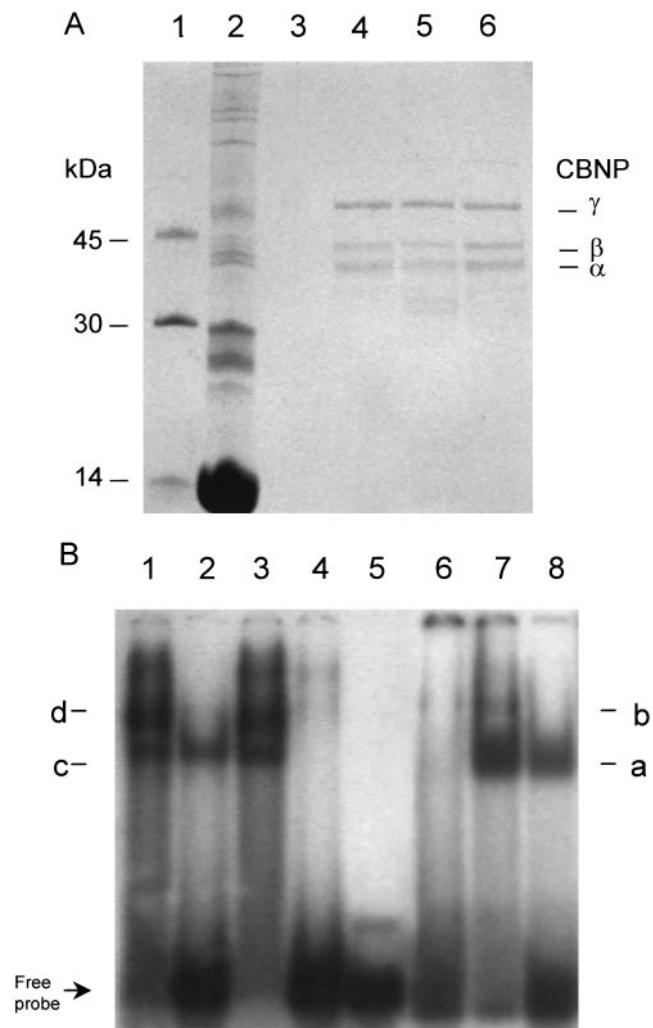
Two software packages, PROWL and PEPTIDESEARCH, were used to identify protein bands from independent non-redundant protein sequence databases which are maintained and updated daily at the European Molecular Biology Laboratory (EMBL) and the National Institute of Health (NIH). They were both used by selection of protein molecular-mass filter, adjustable incomplete cleavage, taxonomic choice of the organism under investigation, peptide mass error, mass changes due to protein modifications (e.g. cysteine alkylation), and isoforms or post-translational modifications. A number of top candidates with high scores from the peptide-matching analysis were further evaluated by comparison with their calculated molecular mass using the experimental values obtained from SDS/PAGE. This parameter was used as a filter with large tolerance (change in molecular mass =  $\pm 15\%$  of the molecular mass) in order to exclude false-positive candidates from the output lists.

### Western-blot analysis

Bands from SDS/PAGE analysis of affinity-purified material were excised from the gel and subjected to further SDS/PAGE in a 10% gel. Proteins were then transferred to nitrocellulose membranes (Schleicher & Schuell, Keene, NH, USA). These were saturated by incubation at 4 °C overnight with 10% non-fat dry milk in NaCl/P<sub>i</sub>/0.1% Tween 20 and then incubated with the rabbit polyclonal anti-[single-stranded D-box-binding factor (ssDBF)] serum for 60 min at room temperature. The rabbit polyclonal anti-ssDBF serum was used at a dilution of 1 : 1000 in NaCl/P<sub>i</sub>/0.1% Tween 20. After three washes with NaCl/P<sub>i</sub>/0.1% Tween 20, they were incubated with an anti-rabbit immunoglobulin coupled to peroxidase (Sigma Chemical Co, St Louis, MO, USA). After 60 min of incubation at room temperature, the membranes were washed several times with NaCl/P<sub>i</sub>/0.1% Tween 20 and the blot was developed using the ECL chemiluminescence method (Amersham Pharmacia Biotech).

## RESULTS

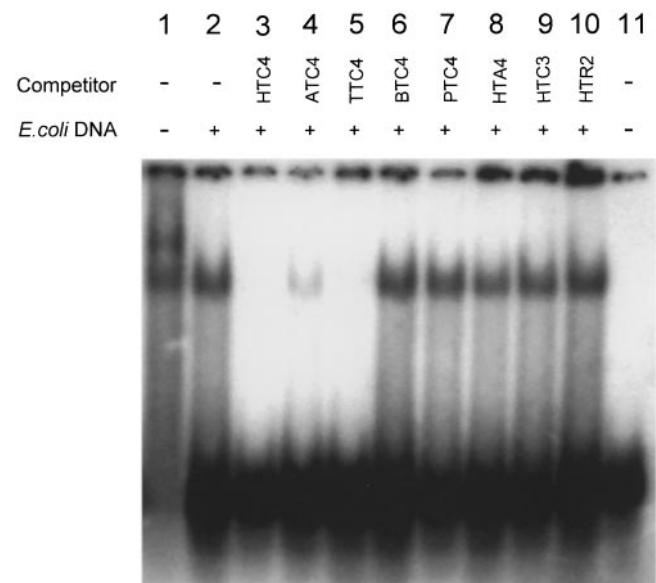
We showed previously in EMSA experiments the presence of proteins that bind to ssDNA with the telomeric repeated motif d(CCCTAA)<sub>n</sub>, but not its cDNA or the duplex, in nuclear extracts from several sources [19,20]. One of these displayed a remarkably high specificity of binding to this sequence. We have now succeeded in isolating these proteins from mature chicken erythrocytes using affinity chromatography.



**Fig. 1.** (A) SDS/PAGE analysis of affinity chromatography recovered components stained with Coomassie Blue and (B) EMSA profiles to track the binding activity toward the labeled probe HTC4 during the isolation procedure. (A) lane 1, molecular-mass markers; lane 2, total extract (10  $\mu$ g) from chicken erythrocytes; lane 3, sample recovered from non-functionalized resin; lanes 4–6, three different affinity recovered samples. The three stained components are indicated in the right margin as CBNP  $\alpha$ ,  $\beta$ , and  $\gamma$ . (B) lanes 1 and 2, whole extract from chicken erythrocytes in the absence and presence, respectively, of 250-fold excess (*w/w*) denatured *E. coli* competitor DNA before incubation with affinity resin. Lanes 3 and 4, whole extract after incubation with non-functionalized and HTC6-functionalized resin, respectively. Lanes 6–8, activity of protein fraction recovered from HTC6-functionalized resin in the absence (lane 6) and presence of 1 mM dithiothreitol [lane 7 in the absence of competitor, lane 8, in the presence of 250-fold excess (*w/w*) denatured *E. coli* competitor DNA]. Lane 5, labeled HTC4 alone.

Three sample cases of SDS/PAGE profiles of the proteins recovered by affinity chromatography are shown in Fig. 1A (lanes 4, 5 and 6), along with the profile of the starting material, i.e. total protein extract (lane 2). Lane 3 shows that no stainable protein is retained by the non-functionalized resin. Essentially three components, with apparent molecular masses of 38, 44 and 55 kDa, were isolated. They have been designated C-block-binding nuclear proteins, CBNP $\alpha$ , CBNP $\beta$  and CBNP $\gamma$ .

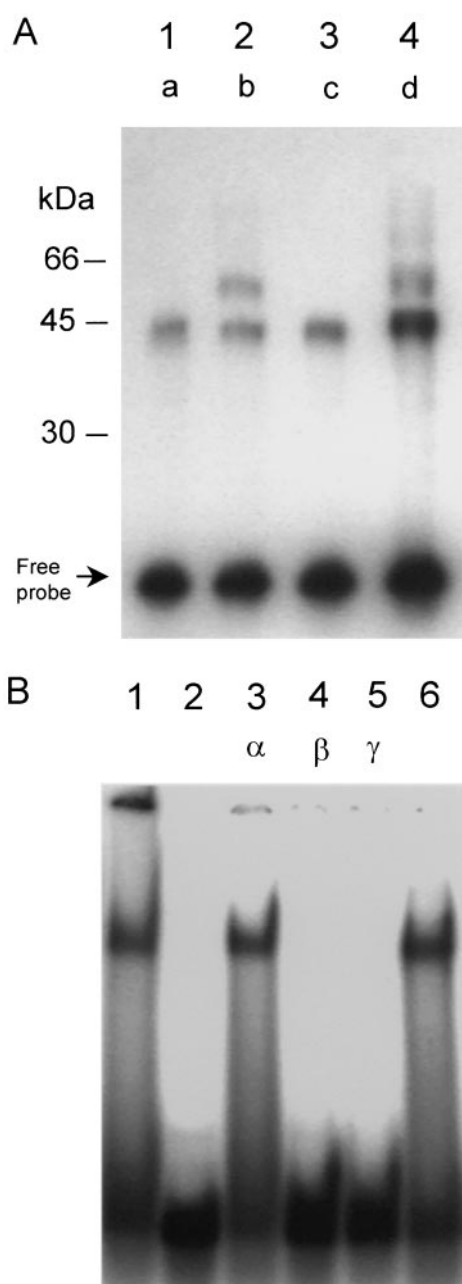
The EMSA profiles of lanes 1–4 of Fig. 1B track the binding activity with the probe HTC4 during the isolation procedure. Lanes 1 and 2 display the activity of the starting material, in the



**Fig. 2.** Competitive EMSA with different telomeric type oligonucleotides. Labeled HTC4 (0.2 pmol) was incubated with 0.1  $\mu$ g affinity recovered proteins in the absence of competitors (lane 1) or in the presence of 250-fold excess (*w/w*) denatured *E. coli* DNA and 10-fold excess of cold telomeric type oligonucleotide competitors (lanes 2–10); lane 11 labeled HTC4 alone.

absence and presence of excess denatured *E. coli* DNA, respectively. After incubation of the extract with the non-functionalized resin, the supernatant retains the binding activity (lane 3), but loses it after incubation with the HTC6-functionalized resin (lane 4). The protein fractions recovered from the functionalized resin, the SDS/PAGE profile of which is shown in Fig. 1A, display the same activity as the whole extract, in both the absence and presence of excess denatured *E. coli* DNA (lanes 7 and 8, respectively). It is important to point out that the presence of dithiothreitol in the incubation buffer was critical for full recovery of DNA-binding activity, as in its absence, no EMSA signal (lane 6) or a smeared signal with low mobility was observed (data not shown).

To test the identity of the DNA-binding activity of the fraction recovered from affinity chromatography with that of the whole extract, a competition EMSA was set up. The formation of the specific protein–HTC4 complex (in the presence of denatured *E. coli* DNA) was challenged with a 10-fold excess of telomeric-related oligonucleotides (Table 1). The result (Fig. 2) matches perfectly with that already seen for whole nuclear extract from HeLa cells [20], despite the different sources. In particular, labeled HTC4 was displaced by unlabeled HTC4 (lane 3) and TTC4 (lane 5) with almost the same efficacy, and slightly less by ATC4 (lane 4), the other sequences being almost completely ineffective (lanes 6–10). *In situ* UV-cross-linking experiments were carried out to establish which of the protein components isolated by affinity chromatography is responsible for each of the EMSA bands, in particular the protein corresponding to the faster and more sequence-specific band. The faster EMSA bands (a and c in lanes 8 and 2 of Fig. 1B) were excised, UV-irradiated, and subjected to SDS/PAGE. In both cases, the protein–probe adduct shows only one signal, the mobility of which is slightly lower than that of CBNP $\alpha$  (lanes 1 and 3 in Fig. 3A) because it is UV-cross-linked with HTC4 which has a molecular mass of  $\approx$  7 kDa. A similar procedure was used to characterize the



**Fig. 3.** (A) SDS/PAGE analysis of *in situ* UV-cross-linking of shifted bands from EMSA gel of Fig. 1B, and (B) EMSA of protein components CBNP $\alpha$ , CBNP $\beta$ , and CBNP $\gamma$  after electroelution of the excised bands from SDS/polyacrylamide gel. (A) Lane 1, UV-cross-linked product of band a in lane 8 of Fig. 1B; lane 2, UV-cross-linked product of band b in lane 7 of Fig. 1B; lane 3, UV-cross-linked product of band c in lane 2 of Fig. 1B; lane 4, UV-cross-linked product of band d in lane 1 of Fig. 1B. (B) Lane 1, whole affinity recovered sample; lane 2, labeled HTC4 probe alone; lane 3, CBNP $\alpha$ ; lane 4, CBNP $\beta$ ; lane 5, CBNP $\gamma$ ; lane 6, the three components electroeluted together.

complexes of slower EMSA bands (b and d of lanes 7 and 1, respectively, of Fig. 1B), in this case two SDS/PAGE signals being obtained, the faster corresponding to the previous one and the other with a mobility slightly lower than that of CBNP $\gamma$  (lanes 2 and 4 of Fig. 3A). This result suggests that the UV-cross-linkable proteins are CBNP $\alpha$  only in the fast EMSA band, and both CBNP $\alpha$  and CBNP $\gamma$  in the slow one. To confirm the identity of the proteins responsible for the specific complex

(fast band), the three components CBNP $\alpha$ , CBNP $\beta$ , and CBNP $\gamma$  were excised separately from the SDS/polyacrylamide gel, electroeluted, freeze-dried, and resuspended in EMSA binding buffer. EMSAs were performed for each of them in the presence of denatured *E. coli* DNA and 1 mM dithiothreitol. It can be seen from Fig. 3B that CBNP $\alpha$  recovered specific DNA-binding activity fully after electroelution (lane 3), consistently with that of the whole fraction obtained from affinity chromatography (lane 1), and after electroelution in the pool (lane 6). On the other hand, CBNP $\beta$  and CBNP $\gamma$ , after separation, did not display any specific EMSA signal (lanes 4 and 5), even in the absence of excess specific competitor (data not shown).

Attempts to determine the nature of CBNP $\alpha$ , CBNP $\beta$ , and CBNP $\gamma$  by direct amino-acid sequencing after electroblotting on poly(vinylidene difluoride) membranes failed because of the presence of an N-terminal blocking group. However, one of the digestion fragments of CBNP $\alpha$  yielded a dodecapeptide MFVGGLSWDTSK. This result was not very instructive, because this sequence corresponds to a conserved RNA-binding motif (RNP1), found in many hnRNPs. Therefore, all the protein species were digested with trypsin *in situ* as described in Materials and methods. The peptide mixtures were extracted and resolved by narrow-bore RP-HPLC for further analysis by MALDI MS. Besides exhibiting similar chromatographic profiles, CBNP $\alpha$  and CBNP $\beta$  yielded many peptides sharing the same mass values (Table 2), suggesting a strong structural relationship between them. These findings were confirmed by independent experiments with endoproteases Asp-N and Glu-C. In fact, CBNP $\alpha$  and CBNP $\beta$  showed very similar peptide electrophoretic profiles (not shown). In contrast, CBNP $\gamma$  gave unrelated peptides in all cases, when digested with trypsin (Table 2) as well as with endoproteases Asp-N and Glu-C (not shown). The molecular-mass values obtained from each of these digestions were used to search protein and nucleotide databanks to identify the nature of each molecular species. Excellent results were obtained in the cases of CBNP $\alpha$  and CBNP $\beta$ , for which 22 out of 24 and 18 out of 20 peptides matched closely those expected for avian ssDBF. This is a 302-residue-long protein and has already been characterized as a liver nuclear factor involved in the transcription of the apoVLDL II gene [21]. The same paper reports the isolation of a further cDNA clone of ssDBF, containing an extra exon, and pointing to the existence of an isoform of this protein. The next best match was obtained for the avian protein CRP1 [22], for which 17 out of 24 and 15 out of 20 peptide matches were observed. Other avian proteins showed fewer than six matches and were not considered. The sequences of ssDBF and CRP1 clearly show that they can be ascribed to the A/B family of hnRNPs. They are identical, except for residues 136 and 221 in the C-terminal region. The N-terminal region of about 80 residues is apparently unrelated.

To clarify further the identity of CBNP $\alpha$  and CBNP $\beta$ , and in particular to discriminate between ssDBF and CRP1, several peptides were subjected to MSMS experiments. In addition to the peptides for which MH<sup>+</sup> signals were observed at *m/z* 1328.5, 1571.9 and 1774.1, corresponding to the expected amino-acid sequences of MFVGGLSWDTSK, EVYQQQFSSGGGR and IFVGGGLNPEATEEKIR, common to both proteins, those detected for MH<sup>+</sup> at *m/z* 1656.0 and 1785.3 yielded fragments with the sequences GFGFILFKEP-GSVEK and GFVFITFKEEDPVKK, respectively (Fig. 4). The presence of Ile136 and Thr221 shows that CBNP $\alpha$  and CBNP $\beta$  are indeed the two isoforms of ssDBF and not of CRP1. In addition, CBNP $\alpha$  and CBNP $\beta$  showed two clear signals at *m/z*

**Table 2.** MS analysis of incognite proteins. CBNP $\alpha$  and CBNP $\beta$  were assigned to two avian ssDBF forms by PROWL search and MSMS experiments. NA, not assigned.

CBNP $\alpha$			CBNP $\beta$			CBNP $\gamma$
Mass	ssDBF	CRP1	Mass	ssDBF	CRP1	Mass
7363.5	Ac2-84	NA	7363.8	Ac(2-84)	NA	1098.7
8108.0	Ac2-91	NA	8107.8	Ac(2-91)	NA	1195.3
1328.5	92-103	75-86	1328.4	92-103	75-86	1505.3
1344.3	92-103ox	75-86ox	1344.5	92-103ox	75-86ox	1519.7
1456.7	92-104	75-87	1456.8	92-104	75-87	1554.4
1472.9	92-104ox	75-87ox	1472.7	92-104ox	75-87ox	1596.1
1159.0	104-112	87-95	1159.2	104-112	87-95	1711.7
1169.9	113-122CAM	96-105CAM	1169.6	113-122CAM	96-105CAM	1818.9
2184.8	113-131CAM	96-114CAM	1656.3	132-146	NA	1834.8
2200.7	113-131oxCAM	96-114oxCAM	1342.7	140-151	123-134	1966.7
3095.2	113-139CAM	NA	2072.5	171-189	154-172	2593.8
3110.9	113-139oxCAM	NA	1505.1	176-189	159-172	2614.7
1656.0	132-146	NA	2485.2	190-210	173-193	2944.8
2110.6	152-169	135-152	2501.1	190-210ox	173-193ox	3084.7
1505.1	176-189	159-172	2215.7	192-210	175-193	3101.1
1774.1	176-191	159-174	2231.4	192-210ox	175-193ox	3832.5
2485.2	190-210	173-193	1941.6	215-230	NA	3895.3
2501.1	190-210ox	173-193ox	1785.3	216-230	NA	3910.9
2215.7	192-210	175-193	1535.2	235-247CAM	218-230CAM	4642.0
2231.4	192-210ox	175-193ox	1571.9	253-266	236-249	
1941.6	215-230	NA				
1785.3	216-230	NA				
1571.9	253-266	236-249				
1334.8	292-302	275-285				

8108.0 and 7363.5 (Table 2) which were tentatively assigned to N-terminal peptides of ssDBF, with the first encoded residue removed and the second one acetylated as the result of co/post-translational processing as previously observed in other hnRNPs.

Further evidence supporting the structural relationship between CBNP $\alpha$  and CBNP $\beta$  has come from ISMS. The protein mixture recovered from the functionalized resin was processed by RP-HPLC to obtain each component in sufficient amounts for MS analysis. SDS/PAGE revealed that the acetonitrile gradient separated CBNP $\gamma$  from the other two components, but not CBNP $\beta$  from CBNP $\alpha$ . This is probably due to their similar hydrophobic profiles. This procedure led to the irreversible loss of the DNA-binding activity. When CBNP $\gamma$  was analysed by ISMS, it did not give any more signals than those corresponding to fragments of about 15 kDa. The spectrum for the CBNP $\alpha$ /CBNP $\beta$  sample was clearly consistent with the presence of two species with molecular masses of  $31915 \pm 2$  Da and  $37718 \pm 4$  Da, respectively, the first being more abundant. The absolute values of their molecular masses were consistently lower than those apparent from SDS/PAGE analysis, but their mass differences, as well as their relative abundance, were in accord with the electrophoretic data. This provided further evidence that these two proteins share several physicochemical properties. Besides the coincidences detected by MALDIMS and MSMS experiments in nearly all peptide segments, the mass difference ( $5803 \pm 4$  Da) between CBNP $\beta$  and CBNP $\alpha$ , determined by ISMS, matches perfectly the value calculated for the extra exon observed in the variant cDNA isolated from the chicken library [21]. Thus, these two proteins can be identified as the two isoforms of ssDBF, differing from each other by the occurrence of an extra exon near the

C-terminus. Both the ISMS molecular masses of CBNP $\alpha$  and CBNP $\beta$  were about 143 Da higher than the values calculated from their sequences. However, it should be noted that they both contain at least two methionine residues in the extensively oxidized form (as determined by MALDIMS mapping) as well as putative methylation sites. Future *in vivo* and *in vitro* studies will elucidate post-translational processing of CBNP $\alpha$  and CBNP $\beta$  and its relationship with their biological activity.

Parallel attempts to search protein and nucleotide databanks using the mass values determined by MALDIMS for the peptides generated from the *in situ* digestion of CBNP $\gamma$  (Table 2) were unsuccessful. This suggests that no protein sequence with high similarity to CBNP $\gamma$  is present in the current databanks. To ascertain the nature of this protein, the purified peptides showing  $MH^+$  at  $m/z$  1098.7 and 1195.3 were subjected to MSMS experiments as previously described. As reported for CBNP $\alpha$ , MSMS analysis of each species yielded fragment ions that allowed us to reconstruct their sequence as GSDFDCELR and NLPLPPPPPPR, respectively. Databank searching revealed that both peptides were identical to two regions of hnRNP K, and were conserved in various species. The corresponding sequence from chicken is still not known. This protein, originally identified as a component of the hnRNP particles, binds to polyC as a consequence of its degree of phosphorylation [23].

Finally, to substantiate further the identity of CBNP $\alpha$  with ssDBF, immunological analysis was performed. With the polyclonal anti-ssDBF serum kindly provided by M. Smidt [21], its ability to recognize the single bands excised from the SDS/polyacrylamide gel of Fig. 1A and the rerun in SDS/PAGE was tested. As can be seen in Fig. 5A, this serum recognized both the bands of CBNP $\alpha$  and CBNP $\beta$ , supporting their identity

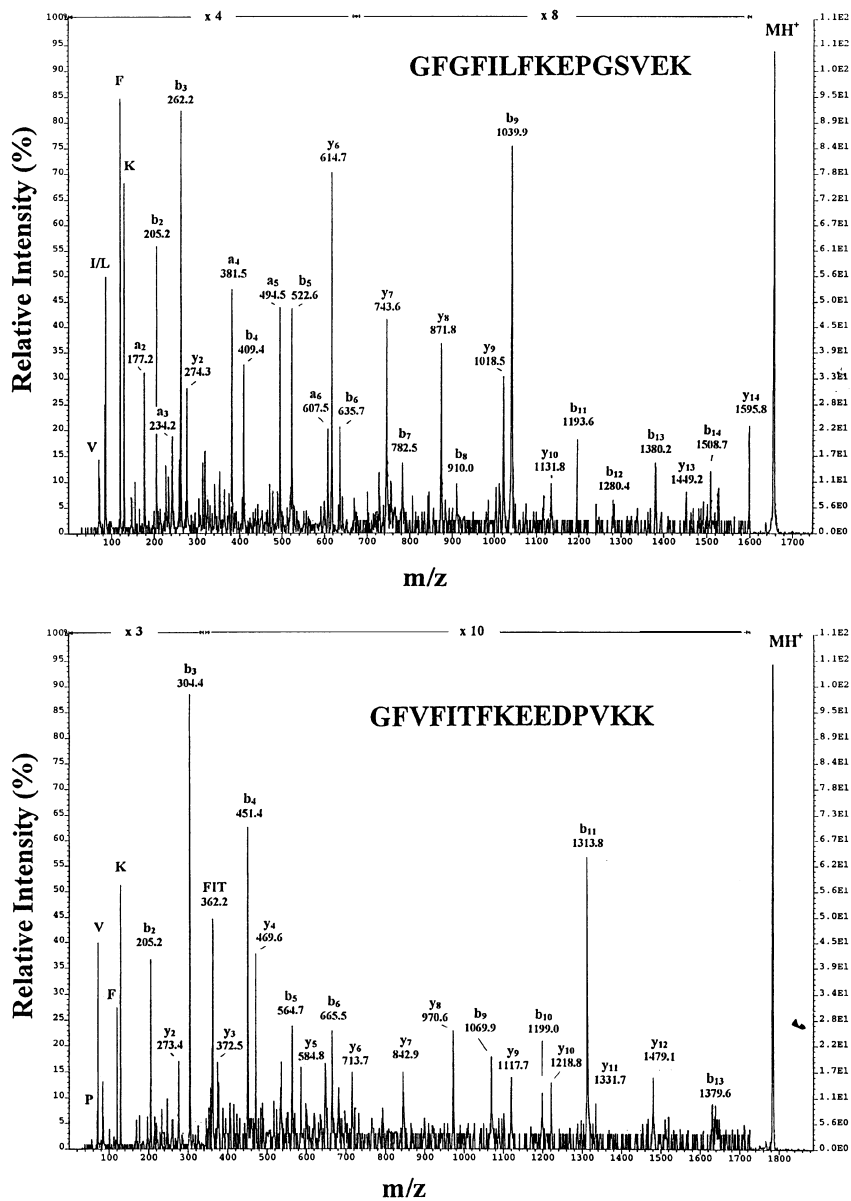


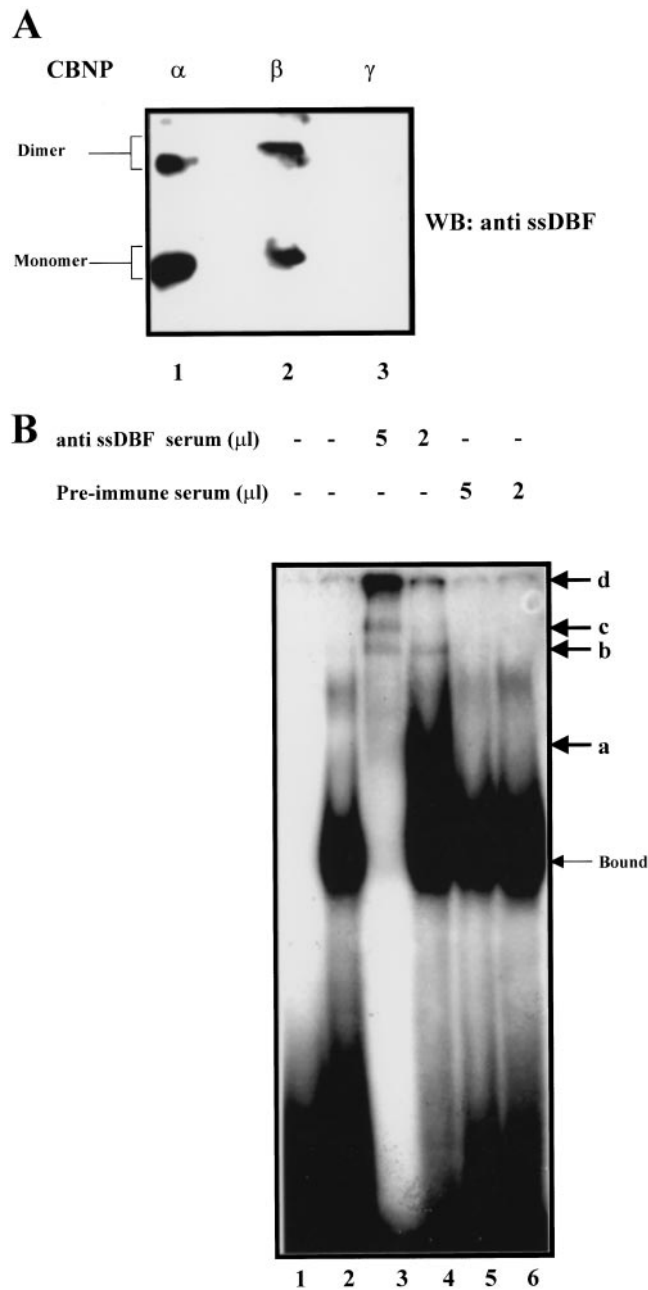
Fig. 4. MS/MS analysis of peptides obtained from in-gel digestion of CBNP $\alpha$ . Identical peptide species were isolated and analysed from the tryptic digest of CBNP $\beta$ .

with ssDBF (CBNP $\alpha$ ) and its isoform (CBNP $\beta$ ). This serum did not recognize the CBNP $\gamma$  band at all, confirming its different nature. The low-mobility band is attributable to the formation of a covalent dimer of the protein, as the gel was run in nonreducing conditions in the absence of 2-mercaptoethanol. Moreover, a supershift analysis with the anti-ssDBF serum was performed on the specific complex of CBNP $\alpha$  with the HTC4 probe. Figure 5B shows that the immune serum is able to supershift the specific EMSA band, giving rise to lower-mobility complexes (lanes 3 and 4, bands a, b, c, d) in a dose-dependent manner. The preimmune serum is not able to supershift any complex at all (lanes 5 and 6). Therefore, together, these immunological data clearly demonstrate the identity of CBNP $\alpha$  and ssDBF.

## DISCUSSION

CBNPs were isolated by affinity chromatography by incubating the extract of chicken erythrocytes and the affinity resin in the absence of a specific ssDNA competitor. With this procedure,

we expected to separate the protein components responsible for all the EMSA signals given by the extract in the absence of competitor. Indeed the EMSA profile of the affinity chromatography fraction was similar to that of the whole extract, displaying both the faster (more specific) and slower (less specific) band in the absence of competitor (Fig. 2, lane 1). Inspection of the EMSA, *in situ* UV-cross-linking, and RP-HPLC experimental results led to the following observations: (a) although CBNP $\beta$  is captured by the affinity resin, it does not UV-cross-link with the probe, and after recovery in the pure form from electroelution it does not give an EMSA signal; (b) CBNP $\gamma$ , which can be UV-cross-linked to the probe in the presence of CBNP $\alpha$ , is found in the slow and less specific EMSA band; (c) during the attempted fractionation of the CBNPs by RP-HPLC, most of the CBNP $\alpha$  and CBNP $\beta$  were coeluted separately from CBNP $\gamma$ , although the latter always retained small amounts of the former two. This may be due to the tendency of these proteins to aggregate through the glycine-rich domains, a well-known feature of other hnRNPs, e.g. hnRNP A1 [24], and to form covalent dimers, as suggested by



**Fig. 5.** (A) Western-blot analysis of the single bands excised from SDS/polyacrylamide gel of Fig. 1A and rerun on a 10% gel, and (B) supershift analysis of the specific complex between CBNP $\alpha$  and  $^{32}$ P-labeled HTC4. (A) Lane 1, CBNP $\alpha$ ; lane 2, CBNP $\beta$ ; lane 3, CBNP $\gamma$ . The bands were assayed by using the specific rabbit polyclonal anti-ssDBF serum provided by M. Smidt [21] and developed by using ECL (Amersham Pharmacia Biotech). WB, Western blot. (B) Samples containing CBNP $\alpha$  and the HTC4 probe were incubated in the absence (lane 2), in the presence of the specific polyclonal rabbit anti-ssDBF serum (lanes 3 and 4) and, as controls, in the presence of the corresponding amounts of preimmune rabbit serum (lanes 5 and 6) for 2 h at room temperature. Then, the samples were analysed by EMSA. The arrows marked a, b, c show the supershifted bands and the d band corresponds to insoluble material precipitated into the well, which, however, is mostly present in the case of the specific serum anti-ssDBF. Lane 1 contained labeled HTC4 alone

the presence of small amounts of species with twice the molecular mass in non-reducing SDS/PAGE (Fig. 5A). The observation that CBNP $\beta$  does not UV-cross-link with HTC4, although it is bound by the affinity matrix, suggests that it

forms a complex with CBNP $\alpha$ , but does not interact directly with the telomeric repeats. As the difference between CBNP $\beta$  and CBNP $\alpha$  is the presence of the extra exon, it can be inferred that this abolishes the DNA-binding activity but not the interaction with other hnRNPs. As far as CBNP $\gamma$  is concerned, the MS data point to its identity with the chicken homolog of hnRNP K. This is in line with the recent work of Lacroix *et al.* [18], who demonstrated that human hnRNP K binds to the single-stranded telomeric C-block motif, although the SDS/PAGE mobility of CBNP $\gamma$  appears to be slightly higher. However, it should be noted that the *Xenopus* homolog, for instance, has a much lower molecular mass than that of human hnRNP K [25]. The observation that the slower EMSA band contains CBNP $\alpha$  besides CBNP $\gamma$  suggests that the two proteins interact, although homotypic protein dimers cannot be excluded. In any case, these complexes are less sequence-specific in their binding to the telomeric repeat probe d(CCCTAA)<sub>n</sub>.

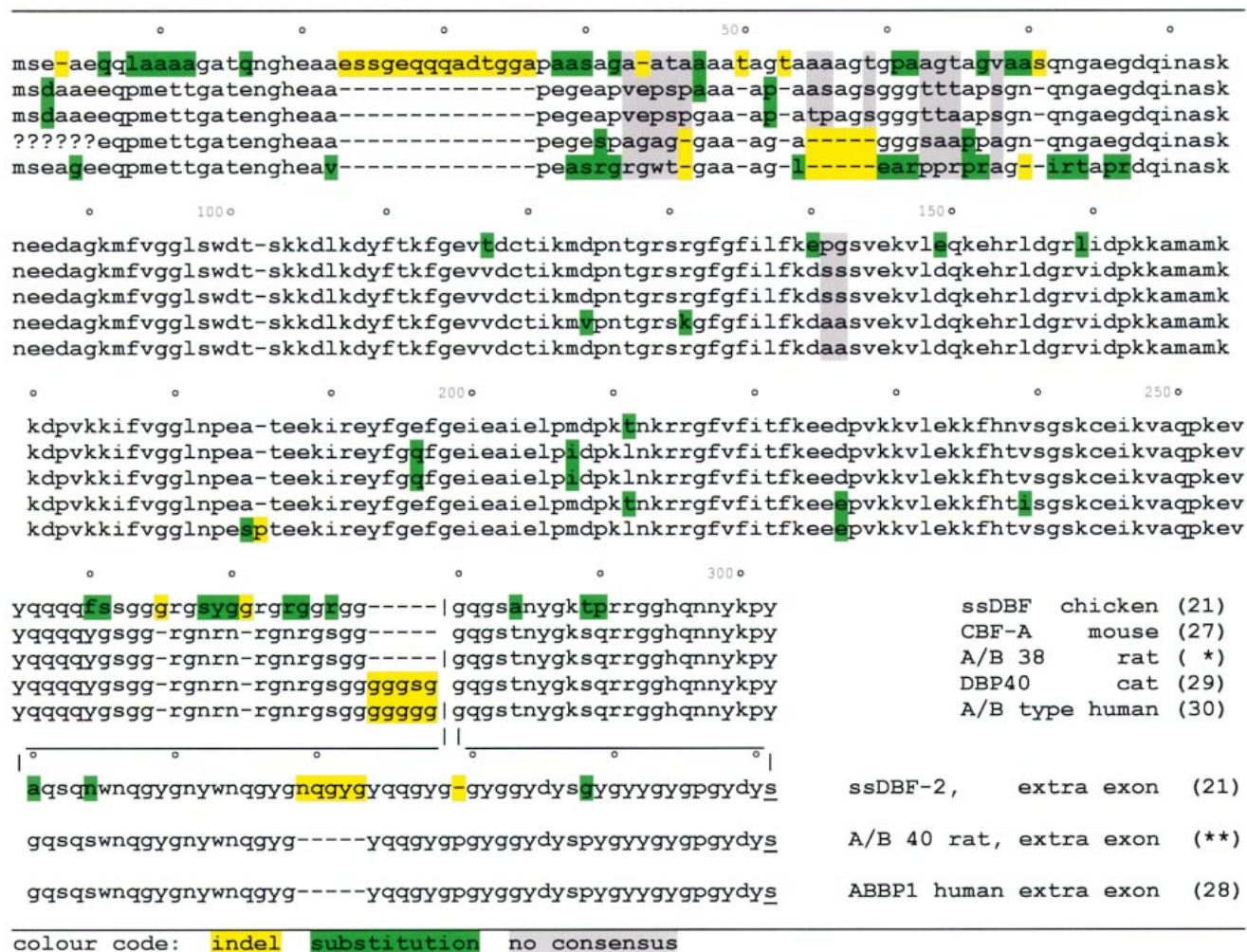
The presence of a reducing agent such as dithiothreitol appears to be required for the specific DNA-binding activity. In its absence, EMSA signals, when found, have low mobility and also lower sequence specificity. Nonetheless, they contain CBNP $\alpha$  as shown by *in situ* UV-cross-linking (data not shown). CBNP $\alpha$  therefore seems to need to be in the reduced state, as well as separated from CBNP $\gamma$ , to be able to specifically interact with the C-rich telomeric repeat.

The identification of CBNP $\alpha$  with the type A/B hnRNP ssDBF raises two issues. The first concerns its possible function in relation to the activity of CBNP $\alpha$ , described here and in two preceding reports [19,20], and that of ssDBF, which has been characterized by Smidt *et al.* [21] as a transcription factor. Strong evidence from studies with cultured cells *in vivo* is required to be able to assign confidently a role to this protein. However, it should be noted that ssDBF was identified during screening of a liver cDNA library, whereas CBNP $\alpha$  was isolated from erythrocytes. It could well be that this protein has more than one role which is tissue dependent, and its activity could be regulated by both the relative amounts of the two isoforms (CBNP $\alpha$  and CBNP $\beta$ , in our notation) and different post-translational modifications.

The second issue is more general and involves how the proteins studied in this work can be related to all similar proteins already known in other species. BLAST analysis [26] carried out using the protein and nucleotide databanks has revealed a set of proteins and/or cDNAs related to CBNP $\alpha$ /ssDBF (Fig. 6). Besides CRP1 from chicken [22], which differs completely from ssDBF in its N-terminal segment as the result of two frameshift mutations in its DNA coding sequence, the other proteins from mouse, rat, cat, and human cDNAs are homologous along their whole lengths, spanning 280–350 residues. The only other exception to this is PRM10 from rat (not shown; GenBank accession No. AF108653), the N-terminal region of which differs completely (also in its DNA coding sequence) from the others. Inspection of these sequences shows the following.

(a) The homology is very strong in the C-terminal portion of all these proteins, covering about 80% of the whole sequence. The 60–70 N-terminal residues of the mammalian sequences (except for PRM10) also display a remarkable similarity among them, to a lesser extent with the chicken protein ssDBF, and, obviously, none at all with CRP1.

(b) In the cases of rat and human sequences, two variants, which differ in the occurrence of an exon, are reported. The proteins with the exon are homologous to CBNP $\beta$  (i.e. the second clone of ssDBF) in sequence and position of the exon.



**Fig. 6.** Alignment of five related hnRNP type A/B sequences from different species. |, Position of the extra exon;  $\xi$  replaces g at the beginning of the last exon in the forms containing the extra exon; ?, unknown residue (probably identical with the corresponding human sequence); \*, S. Leverrier, *et al.* unpublished, GenBank accession No. AJ238855; \*\*, S. Leverrier, *et al.* unpublished, accession No. AJ238854; #, 72 N-terminal residues of CRP1 [22], residues 73–285 identical with residues 90–302 of ssDBF, except for substitutions i to r at position 136 of ssDBF and t to s at position 221 of ssDBF.

Overall, ssDBF appears to be most divergent, in line with its wider phylogenetic gap. Small differences seen in the known sequences (Fig. 6) suggest that these may represent only a subset of a larger group, the outcome of a rather complex series of mutation, gene duplication, and possibly recombination events throughout the evolution of tetrapods.

A functional role has been proposed for some of these proteins (ssDBF as a liver-specific transcription factor [21], CBF-A – and probably the almost identical hnRNP 38 from rat – as a muscle-specific transcription factor [27], and human ABBP1 as an RNA-editing factor [28]) whereas for the feline DBP40 the ability to interact with the 5'-terminal sequence of a panleukopenia virus has been shown [29]. Nothing is known about the functions of the human-type A/B hnRNP [30], CRP1 [22], rat hnRNP 40 and PRM10. In any case, the apparent lack of correspondence between the strong sequence homology displayed by these proteins and their proposed role is intriguing. This varied spectrum of functions may be the consequence of entangled species and

tissue diversification. They may be linked, respectively, to the usual phylogenetic changes and to modulations of their activity, possibly after gene-duplication events, through many, sometimes tissue-specific, isoforms produced by alternative splicing, editing, or post-translational modifications. Maybe these 10 cases from five species, the only ones known so far, are only a small sample of a considerably more extensive family of type A/B hnRNPs yet to be fully unveiled.

## ACKNOWLEDGEMENTS

This work was supported by the Italian Ministry for University and Scientific Research and Technology (MURST), and the Italian National Research Council (CNR). We are indebted to Professor N. Yathindra, University of Madras, India, for helpful discussion. We thank the Central Facility for Mass Spectrometry of the University of Trieste for running the ISMS spectra. We are indebted to Dr Marten Smidt for kindly providing the ssDBF-specific antibody.

## REFERENCES

1. Broccoli, D., Smogorzewska, A., de Chong, L. & Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17**, 231–235.
2. Conrad, M.N., Wright, J.H., Wolf, A.J. & Zakian, V.A. (1990) RAP1 protein interacts with yeast telomeres *in vivo*: overproduction alters telomere structure and decreases chromosome stability. *Cell* **63**, 739–750.
3. Raghuraman, M.K. & Cech, T.R. (1989) Assembly and self-association of *Oxytricha* telomeric nucleoprotein complexes. *Cell* **59**, 719–728.
4. Price, C.M. (1990) Telomere structure in *Euplotes crassus*: characterization of DNA–protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.* **10**, 3421–3431.
5. Schierer, T. & Henderson, E. (1994) A protein from *Tetrahymena thermophila* that specifically binds parallel-stranded G4-DNA. *Biochemistry* **33**, 2240–2246.
6. Sheng, H., Hou, Z., Schierer, T., Dobbs, D.L. & Henderson, E. (1995) Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol. Cell. Biol.* **15**, 1144–1153.
7. Petracek, M.E., Konkel, L.M., Kable, M.L. & Berman, J. (1994) A *Chlamydomonas* protein that binds single-stranded G-strand telomere DNA. *EMBO J.* **13**, 3648–3658.
8. Eid, J.E. & Sollner-Webb, B. (1997) ST-2, a telomere and subtelomere duplex and G-strand binding protein activity in *Trypanosoma brucei*. *J. Biol. Chem.* **272**, 14927–14936.
9. Sarig, G., Weisman-Shomer, P., Erlitzki, R. & Fry, M. (1997) Purification and characterization of qTBP42, a new single-stranded and quadruplex telomeric DNA-binding protein from rat hepatocytes. *J. Biol. Chem.* **272**, 4474–4482.
10. Uchiumi, F., Ohta, T. & Tanuma, S. (1996) Replication factor C recognizes 5'-phosphate ends of telomeres. *Biochem. Biophys. Res. Commun.* **229**, 310–315.
11. McKay, S.J. & Cooke, H. (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGG<sub>n</sub>. *Nucleic Acids Res.* **20**, 6461–6164.
12. Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. & Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115.
13. LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M.R., Wellinger, R.J. & Chabot, B. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet.* **19**, 199–202.
14. Eid, J.E. & Sollner-Webb, B. (1995) ST-1, a 39-kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29-base-pair subtelomeric repeat and its C-rich strand. *Mol. Cell. Biol.* **15**, 389–397.
15. Garcia-Bassets, I., Ortiz-Lombardia, M., Pagans, S., Romero, A., Canals, F., Aviles, F.X. & Azorin, F. (1999) The identification of nuclear proteins that bind the homopyrimidine strand of d(GA-TC)<sub>n</sub> DNA sequences, but not the homopurine strand. *Nucleic Acids Res.* **27**, 3267–3275.
16. Ito, K., Sato, K. & Endo, H. (1994) Cloning and characterization of a single-stranded DNA binding protein that specifically recognizes deoxycytidine stretch. *Nucleic Acids Res.* **22**, 53–58.
17. Tomonaga, T. & Levens, D. (1995) Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.* **270**, 4875–4881.
18. Lacroix, L., Lienard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Helene, C. & Mergny, J.L. (2000) Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres *in vitro*. *Nucleic Acids Res.* **28**, 1564–1575.
19. Marsich, E., Piccini, A., Xodo, L.E. & Manzini, G. (1996) Evidence for a HeLa nuclear protein that binds specifically to the single-stranded d(CCCTAA)<sub>n</sub> telomeric motif. *Nucleic Acids Res.* **24**, 4029–4033.
20. Marsich, E., Xodo, L.E. & Manzini, G. (1998) Widespread presence in mammals and high binding specificity of a nuclear protein that recognises the single-stranded telomeric motif (CCCTAA)<sub>n</sub>. *Eur. J. Biochem.* **258**, 93–99.
21. Smidt, M.P., Russchen, B., Snippe, L., Wijnholds, J. & Ab, G. (1995) Cloning and characterisation of a nuclear, site specific ssDNA binding protein. *Nucleic Acids Res.* **23**, 2389–2395.
22. Cvekl, A., McDermott, J.B. & Piatigorsky, J. (1995) cDNA encoding a chicken protein (CRP1) with homology to hnRNP type A/B. *Biochim. Biophys. Acta* **1261**, 290–292.
23. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289–321.
24. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. & Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* **259**, 337–348.
25. Siomi, H., Matunis, M.J., Michael, W.M. & Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21**, 1193–1198.
26. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
27. Kamada, S. & Miwa, T. (1992) A protein binding to CArG box motifs and to single-stranded DNA functions as a transcriptional repressor. *Gene* **119**, 229–236.
28. Lau, P.P., Zhu, H.J., Nakamuta, M. & Chan, L. (1997) Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing. *J. Biol. Chem.* **272**, 1452–1455.
29. Wang, D. & Parrish, C.R. (1999) A heterogeneous nuclear ribonucleoprotein A/B-related protein binds to single-stranded DNA near the 5' end or within the genome of feline parvovirus and can modify virus replication. *J. Virol.* **73**, 7761–7768.
30. Khan, F.A., Jaiswal, A.K. & Szer, W. (1991) Cloning and sequence analysis of a human type A/B hnRNP protein. *FEBS Lett.* **290**, 159–161.