Nucleic Acids Research, 2001, Vol. 29, No. 11 ***2427–2436***

Selection and identification of proteins bound to DNA triple-helical structures by combination of 2D-electrophoresis and MALDI-TOF mass spectrometry

F. Guillonneau, A. L. Guieysse, J. P. Le Caer[1], J. Rossier[1] and D. Praseuth*

Laboratoire de Biophysique, INSERM U201, CNRS UMR 8646, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris Cedex 05, France and [1]Laboratoire de Neurobiologie et diversité cellulaires, CNRS UMR 7637, E.S.P.C.I, 10 rue Vauquelin, 75231 Paris Cedex 05, France

Received November 9, 2000; Revised February 2, 2001; Accepted April 5, 2001

ABSTRACT

Identification of proteins binding specifically to peculiar nucleic acid structures can lead to comprehension of their role *in vivo* and contribute to the discovery of structure-related gene regulation. This work was devoted to establishing a reliable procedure to select proteins on the basis of their interaction with a nucleic acid probe chosen to fold into a given structure. 2D-electrophoresis and mass spectrometry were combined for protein identification. We applied this procedure to select and identify triplex-binding activities in HeLa nuclear extracts. To achieve this, we used a panel of deoxyribonucleic probes adopting intramolecular triple-helices, varying in their primary sequence, structure or triple-helix motif. A limited number of spots was reproducibly revealed by South-western blotting. Spots of interest were localised among a complex population of ^{35}S-labelled proteins according to their ^{32}P-specific emission. Position of the same spots was extrapolated on a preparative gel coloured with Coomassie blue, allowing excision and purification of the corresponding proteins. The material was subjected to mass spectrometry upon trypsin digestion and MALDI-TOF peptide fingerprinting was used for research in databases: five of them were identified and found to belong to the hnRNP family (K, L, A2/B1, E1 and I). The identities of several of them were confirmed by comparing western and South-western blots on the same membrane using specific antibodies. The recognition specificity of most of these proteins is large, according to previous reports and our own experiments. It includes pyrimidine-rich DNA sequences in different contexts: single strand to a small extent, triplex and possibly other higher-order structures.

INTRODUCTION

DNA can adopt peculiar non-B structures dictated (i) by the primary nucleotidic sequence and (ii) by environmental parameters such as ionic strength and presence of specific ligands. For example, H-DNA designates a structure that can be adopted by mirror-repeat oligopurine.oligopyrimidine sequences (1); folding back of one of the DNA strands to establish additional Hoogsteen type hydrogen bonds with the purine strand of the duplex results in triplex formation, whereas the complementary strand remains unpaired. The possibility of self-contact between two mirror sequences separated by a long distance has suggested that H-DNA could also have a role in chromatin structuration and condensation (2). Antibodies specific for triple-helical DNA were shown to interfere with the global transcriptional activity as shown by run off experiments performed on isolated nuclei (3). However, the biological function of these structures remains difficult to ascertain.

Deciphering the role of H-DNA requires the search of intracellular partners able to specifically recognise and stabilise these structures. Although the role of H-DNA structure in gene expression was suspected in a panel of known genes, a general mechanism of regulation has not yet been demonstrated. No example has so far established the role of protein binding specifically to the three-stranded part of DNA, although a few reports have dealt with such proteins (4,5). In HeLa cell extracts evidence for such an activity was provided by gel retardation of poly(dA·dT)2 (6), of intramolecular triplex formed with a motif found in IL2Rα promoter sequence (5) or of covalent triplexes formed with psoralen derivatives (4). These proteins were characterised according to their apparent molecular weight in SDS–PAGE without further identification. Moreover, identification of these proteins can be of high interest with regard to the antigene strategy, whereby short oligonucleotides (ODN) are used to artificially regulate gene expression via triplex formation on their genomic DNA target.

In this work, we developed a new method for systematic investigation of proteins that would specifically recognise peculiar nucleic acid structures. To search for triplex-binding proteins, 2D-electrophoresis was combined with South-western blotting to reveal proteins binding specifically to a 55mer DNA probe previously shown to adopt an

*To whom correspondence should be addressed. Tel: +33 1 40 79 37 08; Fax: +33 1 40 79 37 05; Email: praseuth@mnhn.fr

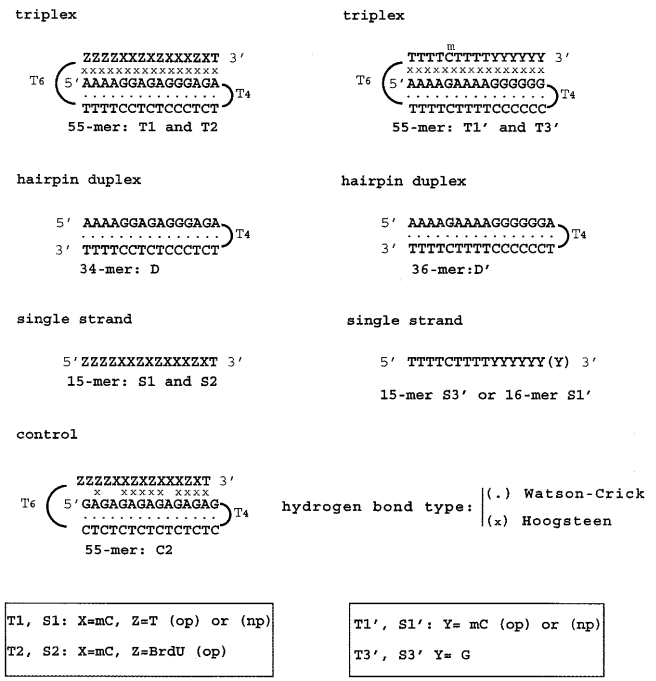

Figure 1. Sequences and structures of the DNA probes. 55mers adopting intramolecular triplex structures are represented. The letters **T**, **C**, **D**, **S** stand for Triplex, related Controls of less stability, Duplex and Single strand, respectively. Indices (1, 2, 3) stand for different nucleotidic compositions in the third strand. (**T**, **D**, **S**) and (**T′**, **D′**, **S′**) are used for distinct duplex motifs. (op) and (np) stand for the nature of phosphodiester and phosphoramidate backbone, respectively. A variation was introduced for **T1** probe, replacing T4 and T6 loops by tri-ethylene glycol **T1**(teg).

intramolecular triplex structure (Fig. 1). We took advantage of techniques recently developed for proteome analysis to establish a correspondence between proteins and the genome (7–9). Proteins physically separated by 2D-electrophoresis can be excised, submitted to trypsin digestion and the resulting peptides extracted from a spot on acrylamide gel, and analysed by mass spectrometry. The method originally set out for the search of triplex or H-DNA binding proteins can be adapted for the investigation of any protein specifically recognising either a given primary nucleotidic sequence, or other higher-order DNA structures. Comparison of hybridisation patterns obtained with appropriate nucleic acid probes used as controls can lead to the selection of proteins potentially harbouring a structural specificity. This method applied to the screening of human HeLa nuclear extracts led to the identification of several of the spots revealed by DNA probes adopting triplex structures. Confirmation of some of them was obtained by western blot performed on the same membrane, using specific antibodies when available.

MATERIALS AND METHODS

Chemicals

ODN with a phosphodiester (op) backbone were provided by Eurogentech and purified on a 15% 19:1 acrylamide:bisacrylamide (Quantum Bioprobe) denaturing gel. ODN with N3′→P5′ phosphoramidate (np) backbone were obtained from Lynx.

Cell culture and metabolic labelling

HeLa cells were obtained from ECAC and grown in 7% FCS (Dutscher), 0.2 mg/ml streptomycin, 200 U/ml penicillin and DMEM (Sigma). To obtain radiolabelled cell extracts, the cells were grown in 15 cm Petri dishes (Corning) to reach 60% confluence, washed twice in PBS and methionine depleted with methionine-free DMEM for 2 h. The medium was then supplemented with 13.5 µCi/ml of [^{35}S]methionine (Amersham Pharmacia Biotech) for 3 h.

Cell-free nuclear extracts

Homemade HeLa extracts were prepared according to the Dignam protocol (10) with the following modifications for the composition of lysis solution: 1 mM sodium orthovanadate (Sigma), 1 mM sodium pyrophosphate (Sigma), 125 nM okadaïc acid (Sigma), 1 mM EGTA (Merck), 0.2% NP-40 (Sigma). HeLa cell nuclear extracts were also available commercially from Promega 'Helascribe nuclear extract' (*in vitro* transcription grade ref. E3092).

1D-gel electrophoresis

HeLa nuclear extracts (20 µg/cm) were loaded on the single 20 cm wide well of a 12% reducing SDS–PAGE [according to Laemmli (11)]. Samples were loaded without prior heating but equilibrated in the loading buffer for 3 min at room temperature. For western and South-western blots, the gel was electrotransferred onto a nitrocellulose membrane (RPN 303 from Amersham).

2D-gel electrophoresis

IsoElectroFocusing drystrip kits were purchased from Amersham Pharmacia Biotech and used as indicated in the manual. Nuclear extracts (150–500 µg) were loaded, and after the first dimension, the strips (13 cm) were equilibrated for 15 min in 50 mM Tris–HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 60 mM DTT. After equilibration, the strips were loaded on a 1 mm thick 10% acrylamide:bisacrylamide (37.5:1, w:w) vertical SDS–PAGE gel, along with molecular weight markers (broad range from Amersham Pharmacia Biotech). Gels were run at 130 V overnight and were either blotted or stained.

Electroblotting

The gels were transferred onto nitrocellulose membranes using a transblot semi-dry apparatus system from Amersham Pharmacia Biotech, the transfer buffer was made of 10% methanol, 0.05% SDS, 5.81 g Tris base and 2.93 g glycine per litre. Membranes were stained with 0.5% Ponceau S in 1% aqueous acetic acid for 1 min and destained with deionised water.

Gel staining

Immediately after the second dimension run, the gels were soaked in a 0.1% solution of Coomassie blue R250 in 35% methanol and 10% acetic acid for 1 h on a rocking shaker. Destaining was performed with the same solution without dye. Once destained, the gels were washed several times with bi-distillated water, and kept wet in 1% acetic acid.

South-western blotting

Transferred proteins were left for at least 1 h in hybridisation solution (50 mM HEPES pH 7, 5 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.05% Tween-20) on a rocking shaker at 4°C. Hybridisation was performed after addition of single- and double-stranded DNA competitors (10 µg/ml dT26 Eurogentech, 30 µg/ml sonicated Herring sperm DNA from Roche) and the ^{32}P-labelled DNA probe (0.2 nM). The total activity (c.p.m.) and probe concentration was kept constant from one experiment to another. The final volume of hybridisation solution was 10 or 2 ml for 2D- or 1D-membranes, respectively. Hybridisation was performed overnight at 4°C. The membrane was rinsed twice for 10 min in hybridisation solution without DNA and briefly dried on a Whatmann paper. The blot was then exposed between two Kodak Molecular Dynamics phosphor screens, with an aluminium sheet intercalated on one side to select ^{32}P radioactivity while masking ^{35}S when necessary.

Quantification of signal intensity obtained from South-western blot

The signal intensity was measured using ImageQuant 5 software. Each spot was surrounded by a circular or elliptic object and the signal measured in volume, the corresponding background (b) was substracted ($s - b$). To take into account the possible variation of the signal that could for example be ascribed to different exposition time, this value was divided again by (b). The value $(s - b)/b$ is reported in Table 1.

Correlation between 1D- and 2D-data

The correspondence between 1D bands and 2D spots was made possible by loading a small amount of extract in an adjacent well before running in the second dimension.

Antibodies. B4B6 (monoclonal provided by Dr J.E.Celis) and 12G4 (monoclonal provided by Dr G.Dreyfuss) both recognise hnRNP K (12). Anti-hnRNP (A2 and I) was provided by Dr G.Dreyfuss (13), Anti-hnRNP E1 and E2 were provided by Dr A.Andino (14). Anti p54nrb antibodies was provided by Dr A.R.Krainer.

Western blots. Western blots were performed according to the procedure described by Lacroix *et al.* (15).

MALDI-TOF analysis

The acrylamide spots were submitted to acetonitrile washing, and reductive alkylation by iodoacetamide in ammonium carbonate (0.1 M) (30 min in dark). In-gel trypsin digestion (Roche, EC 3.421.4) was allowed as described by Shevchenko *et al.* (16). Peptides were resuspended in 20 µl formic acid 1%, desalted using Zip Tip C-18 (Millipore) and eluted with 50 and 80% acetonitrile. The desalted peptide mixture was dried and dissolved in 3 µl formic acid 1%. The matrix used was a saturated solution of 2,5-dihydroxybenzoic acid in TFA 0.1%. The sample and the matrix (1:1, v/v) were loaded on the target using the dried droplet method. MALDI-TOF spectra of the peptides were obtained with a Voyager-DE STR Biospectrometry Workstation mass spectrometer (PE Biosystems Inc.). The analyses were performed in positive ion reflector mode, with an accelerating voltage of 20 000 V, a delayed extraction of 200 ns and around 250 scans were averaged. For subsequent data processing, the Data Explorer software (PE Biosystems Inc.) was used. Spectra obtained for the whole protein were calibrated externally using the $[M+H]^+$ ion from Des-Arg Bradykin peptide (mol. wt 904.4681 Da) and ACTH peptide (mol. wt 2465.1989 Da). The trypsin auto-proteolysis products (132–142 fragment, mol. wt 1153.57 Da and 56–75 fragment, mol. wt 2163.06 Da) were used as the external calibration standard. A mass deviation of 0.1 Da was allowed in the database searches.

Data bank search

The measured peptide masses were used as inputs to search the NCBI or SWISS-PROT databases, using the PROFOUND, EXPASY's peptident and MSFIT search engines. The allowed molecular mass range was ±15% of the experimental value determined with the position of the spots compared with molecular weight markers. Isoelectric points were allowed to range from ±1 pH unit around the experimental value determined by its position according to the first dimension. Oxidation of methionine was included as a possible side reaction. The reported results were selected on the basis of restriction to species *Homo sapiens*, with a maximum number of two missed cleavages, scores of at least six matching peptides among the most intense masses measured and 20% of covered sequence were selected in priority (7).

RESULTS

We used different probes to screen proteins specific for triplexes presented in Figure 1 and to determine the contribution of different parameters in the recognition by specific proteins such as (i) the target sequence (the prime indicates a second duplex sequence), (ii) the type of third strand motif (1, 2, 3 specify different motifs) and (iii) the chemical nature of the third strand (backbone op for phosphodiester derivatives or np for phosphoramidates N3′→P5′). The letters **T**, **C**, **D**, **S** stand for Triplex, related Controls of less stability, Duplex and Single strand, respectively. The comparison of these probes allowed us to evaluate the influence of triplex stability on the formation of complexes and the relative part of target or third strand sequence.

A specific complex revealed by gel retardation with 55 nt intramolecular triplexes in the presence of nuclear HeLa extracts

Detection of a triplex structure specific complex has already been described (5). Conditions were found under which one complex was specifically observed with the triplex form (55mer) but not with the corresponding duplex or third strand. A common retarded band (5) was observed using intramolecular triplexes with different sequences: **T1** and **T2**, with a third strand containing cytosines and either thymines or bromodeoxyuracils (CT or CBrdU motifs), **T3′** with a third strand composed of guanines and thymines except one cytosine (TCG or G-rich motif) (see Fig. 1). In some experiments, it was observed that a small amount of the probe was trapped in a second complex specific for the duplex part, probably reflecting the existing equilibrium between the opened and closed triplex forms (these forms are so named according to the state of the third strand, either dissociated from or Hoogsteen

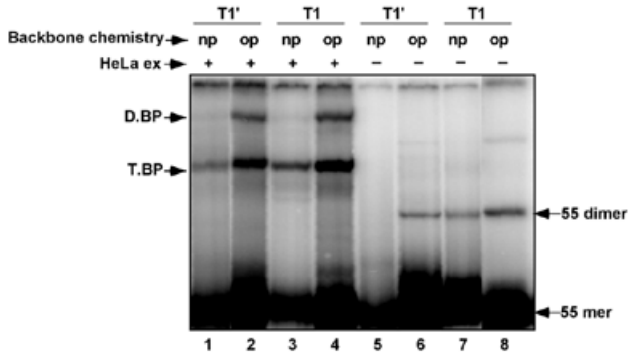

Figure 2. Gel retardation with 55mer intramolecular triplexes. Lanes 1–4: in the presence of 2.5 µg of nuclear HeLa extract; lanes 5–8: without extract. Lanes 1, 2, 5 and 6: **T1′**; lanes 3, 4, 7 and 8: **T1** lanes 1, 3, 5 and 7: the third strand has a np backbone; lanes 2, 4, 6 and 8: the third strand is op. D.BP and T.BP stand for duplex and triplex-binding proteins, respectively, as previously shown by competition experiments (5). 55mer designs a bi-molecular species where the hairpins are probably paired.

base paired to the duplex. To circumvent this problem we decided to further increase the triplex stability by replacing the op backbone with a N3′→P5′ np backbone. This type of modification was already shown to stabilise intermolecular triplexes in the Hoogsteen motif, especially at neutral pH in the case of the (TC) motif (17). As expected, a similar retardation pattern was observed for the derivatives **T1** (op) or **T1′** (op) (referred to simply as **T1** or **T1′**), and **T1** (np) or **T1′** (np) (Fig. 2), except that the protein specific for the duplex part (lanes 2 and 4) was faintly detected for **T1**(np) and **T1′**(np) (lanes 1 and 3), reflecting an equilibrium shift in favour of the triplex form.

Selecting triplex-binding proteins through South-western blot analysis

1D-analysis: a profile of bands specific for triplex structure? To attempt to identify the protein seen by gel retardation, experiments were performed in which bands revealed by the triplex probe were detected by direct formation of the complex on a blotted protein gel. This South-western method was previously described by others (12). A highly reproducible pattern in the presence of the **T1** or **T2** probes revealed a profile of discrete radioactive bands (Fig. 3) emerging from the general ^{35}S-labelled protein pattern depicted in lane c. These bands will be designated by bold arabic letters chosen as a function of the name of the corresponding protein(s) whenever identified. We compared this pattern with the one obtained with control probes constituting for one part of the duplex **D**, and for the other part of the third strand alone **S**. A very faint signal was observed using **D** as a probe (not shown), under the conditions used for membrane incubation (large excess of single-stranded and duplex competitors). A second probe **T′** harbouring a different target sequence was tested in these experiments and was available with several third strand motifs (TC or TCG for **T1′** and **T3′**, respectively). Comparing the probe **T3′** (Fig. 3, lane 9) and the corresponding control **S3′** (lane 16), bands were observed with higher intensities for the triplex compared with the 15mer: this stands for band **kl** (66–68 kDa) and **w** (75 kDa). A few bands were observed only with the triplex (not with the control **S3′**), for example **ey** (40 kDa). Conversely, many bands were only observed with the control **S3′** (Fig. 3, lane 16). Some of the proteins such as both **kl** (66–68 kDa) and **ijp** (60–64 kDa) were and observed with all triplexes (Fig. 3, lanes 3, 5 or 12) and their respective control third strand **S1**(np) and (op) (Fig. 3, lanes 4 and 15). For probes containing the pyrimidine third strand motif, signals obtained with the triplexes were always more intense than with the corresponding 15mer. This was clearly seen when comparing **T1**(np) (Fig. 3, lanes 3 and 5) to **S1**(np) (Fig. 3, lane 4), or **T1′** (Fig. 3, lane 12) to **S1′** (Fig. 3, lane 15). Competition studies performed with the ^{32}P-labelled **T2** (or **T3′**) probe in the presence of 100-fold triplex or single strand concentration showed that **S2** (or **S3′**) did not compete with complex formation whereas substantial competition was observed with **T2** (or **T3′**) (Fig. 3, lanes 8 and 11 compared with lanes 7 and 10). This indicated that an important fraction of the 55mer actually adopted a completely folded conformation probably inaccessible to single-stranded or duplex-binding protein. The data suggest that elements other than opened third strand and 34mer hairpin play a role in the formation of complexes.

When comparing the pattern obtained with two triplex probes harbouring different motifs for the third strand, it appeared that **T3′** revealed a clearly different pattern as compared with **T1′**. Several common bands (**x**, **kl**, **ijp**, **z1**) were observed with both probes, but additional specific bands (**w**, **ey**, **z2**) were seen in the case of **T3′** (Fig. 3, lane 9) and did not appear with **T1′** (Fig. 3, lane 12). **x** is probably not specific for triplexes as it can be seen in every lane.

Remarkably, the profiles obtained with **T1′** (Fig. 3, lane 12) and **T1** and **T2** (Fig. 3, lanes 1, 2, 3 and 5), were more alike than the one obtained with **T3′** (Fig. 3, lane 9). It could be concluded that the third strand motif, either composed of pyrimidine (**T1**, **T1′**, **T2**) or guanine-rich (**T3′**), is more important than the duplex sequence in structure recognition, at least in this case. This point was further confirmed by 2D-electrophoresis.

The chemical nature of the third strand backbone did not qualitatively change the profile, but some of the bands (**kl**, **ijp**) had enhanced intensity with **T1**(np) (Fig. 3, lanes 3 and 5) as compared with **T1**(op) (Fig. 3, lane 1). Intermolecular triplexes formed at pH 7 between **S1**(np) and **D** one of which was labelled (Fig. 3, lanes 17–19) revealed mainly band **ijp**.

2D-electrophoresis analysis: a pattern of spots specific for triplex structure? In order to increase the resolution of protein detection and to progress further in protein identification, 2D-electrophoresis was performed to separate proteins from HeLa nuclear extracts, according to the method of immobilised gradients (IPG) (see Materials and Methods). A wide pH range (pH 3–10) was initially chosen because we reasoned that restricting the research to basic proteins could favour the selection of DNA-binding proteins involving mainly electrostatic interactions and devoid of specificity. Furthermore, nucleic acid binding proteins already described in the literature could be found in a large range of isoelectric points.

Experimental conditions were first established to obtain a reproducible pattern of spots. Figure 4 shows the comparison of images obtained by ^{35}S (Fig. 4A) or Coomassie detection (Fig. 4B) and ^{32}P detection (Fig. 4C–E). A limited number of

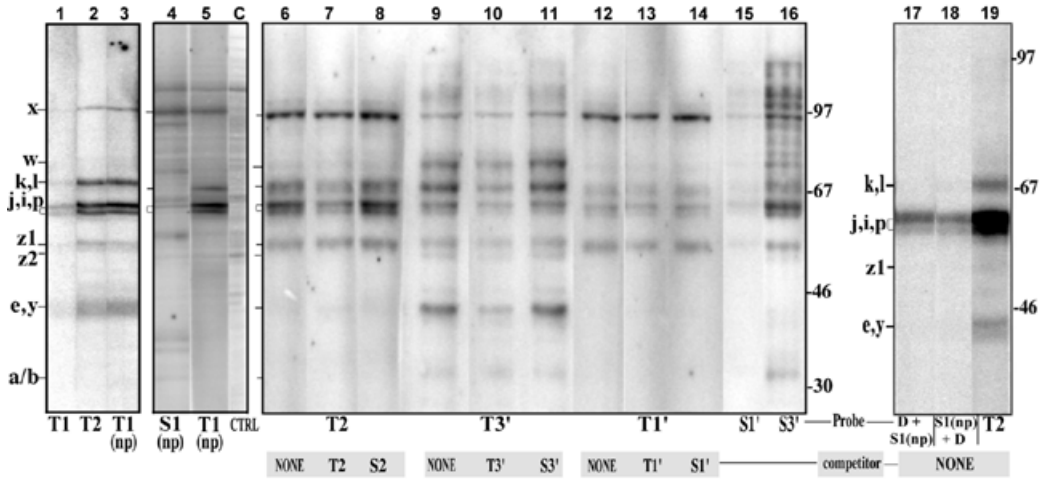

Figure 3. South-western blot of 1D-protein gels in the presence of intramolecular or intermolecular triplexes. Intramolecular 55mers were 5′-labelled with **T1** and **T2** (op) containing either a op third strand (lanes 1 and 2, respectively), or a np third strand **T1**(np) (lanes 3 and 5). Lanes 3, 4 and C were made with radiolabelled nuclear extracts, lane C was not hybridised. The control formed with labelled **S1**(np) is shown in lane 4. Experiments were made in the presence of different third strand motifs and the corresponding unlabelled competitors: **T2**, (CBrdU): lanes 6, 7 and 8; **T3′**, (TCG) lanes 9, 10 and 11; **T1′** (TmC) lanes 12, 13 and 14, and controls obtained with the third strand alone (**S3′**: lane 16) and **S1** (lane 15). Competition experiments were performed in the presence of a 100-fold excess (20 μM) of cold **T** (lanes 7, 10 and 13), or cold **S** (lanes 8, 11 and 14). South-western blot was performed with intermolecular triplexes formed between **S1** (np) and 5′-labelled **D** or 5′-labelled **S1** (np) and **D** and was compared with intramolecular triplex **T2** (lanes 17, 18 and 19, respectively). The proteins are designed by bold arabic letters chosen as a function of the name of the corresponding protein(s) whenever identified. The different windows correspond to different gel migration experiments.

spots emerged as highly labelled from the profiles obtained upon South-western blotting, using **T2** and **T1**(np) (Fig. 4C and E). As the correspondence between some radioactive bands (1D) and spots (2D) could be established (Materials and Methods), the same annotation of spots by bold arabic letters was used. No signal was seen with **D**. Some spots were also revealed by **S2** (for example spots **k**, **e**; Fig. 4D) although with a lower intensity; this pattern was found to be reproducible. Table 1 reports the relative signal intensity for the different spots corresponding to each probe depicted in Figure 1. In addition, the relative abundance of each species is reported, using its associated ^{35}S activity. To ensure that an opened 55mer could not fully account for complex formation, we used the 55mer control **C2** (Fig. 1) in which the sequence of the hairpin had been changed to an alternating AG, while maintaining an identical third strand. The pattern of **C2** and **S2** was similar (Table 1), suggesting that the length of the oligonucleotide does not play a crucial role. Experiments performed with **T1** and **T1′** gave exactly the same pattern as **T2** except that the intensities of the spots were generally lower (Table 1). This could reflect the higher triplex stability of the **T2** derivative (5).

South-western experiments were again performed with **T3′** to evaluate the influence of the third strand motif. Many spots were common for both triplex probes, but their relative intensities were different (see Table 1). Furthermore, some spots appeared more specific for the TCG triplex motif (**y**, **w**), whereas others (**k**, **e**) were preferentially seen with the (CT) or (CBrdU) motif. These results corroborated 1D studies: it was observed that band **ey** was always revealed with **T3′**, whereas **e** could be lost with **T1**, **T2** or **T2′** according to its higher sensitivity towards denaturation conditions.

Replacing the op by the np backbone in the (TC) motif did not significantly modify the pattern (Table 1). Conversely, a significant difference was observed for single-stranded (op) and (np) ODN corresponding to control third strand: **S1**(np) revealed multiple proteins with weak intensities (not shown), while 15(op) **S1** or **S2** (Fig. 4D) revealed a limited number of spots.

To evaluate the possible role of thymine-loops in the recognition process, we repeated the same experiment using a 45mer probe that contained the same nucleotide sequences as **T1** except that the 15 nt motifs were connected with triethyleneglycol **T1** (teg) linkages instead of the two linkers composed of six and four thymines found in **T1**. Loops could be recognised by proteins having an affinity for single-stranded pyrimidine-rich DNA or RNA. **T1** (teg) gave a pattern where spots **k**, **l**, **e**, **i** were clearly detected. It could, thus, be concluded that a majority of the spots did not involve binding to (T)n loops. Intermolecular triplexes formed with **S1**(np) and **D** were also tested and in addition to protein **i** detected in 1D-analysis, spots **k**, **e** and **l** were clearly detected probably because the input protein amount was higher (not shown).

In summary, the strongest signals obtained in the presence of the most stable triplex probes **T2** and **T**(np) derivative (Table 1) revealed **k**, **j**, **e** and **i** whereas **ab** was less visible in spite of its abundance (see Table 1).

Selecting spots revealed by triplex probes for protein identification

Detection of spots was performed by South-western annealing of the blotted 2D-protein gel with various nucleic acid probes. Some of the spots seen in Coomassie blue-stained gels were excised, and subjected to trypsin digestion; the peptides were then extracted from acrylamide (see Materials and Methods), and the material was submitted to MALDI-TOF giving a peptide mass fingerprint. This operation unavoidably required the comparison of two 2D-gels: one for South-western analysis,

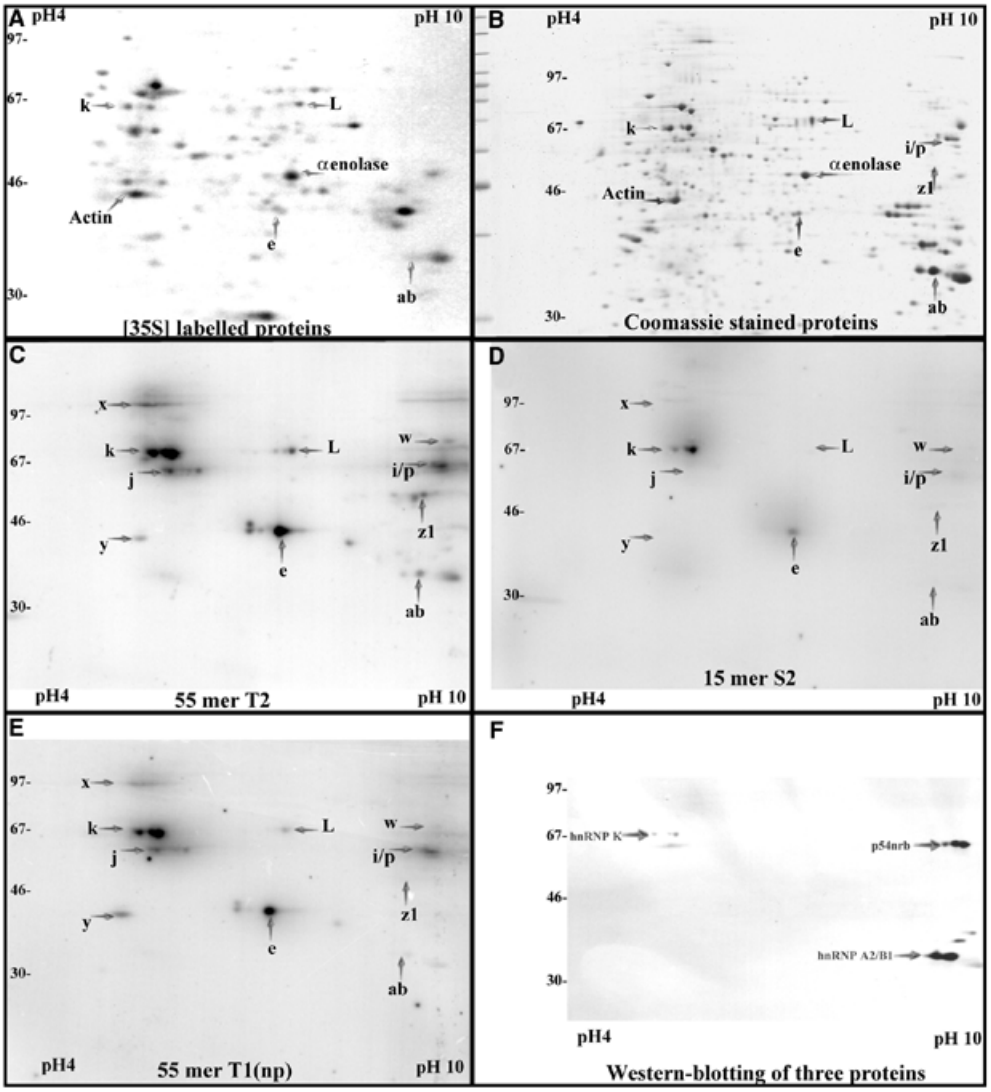

Figure 4. (**A**) and (**B**) show the pattern of ^{35}S-labelled (A) or Coomassie-labelled (B) proteins of nuclear extracts. The spots of interest are indicated by arrows. (**C**) and (**E**) are South-western blots of 2D-gels performed with intramolecular 55mer triplexes. The third strand in the polypyrimidine motif, is composed of CBrdU (C, **T2**), or CT [E, **T1**(np)]. The control corresponding to annealing of the third strand alone are shown in (**D**) (**S2**). Western-blot performed on the same membrane in the presence of specific antibodies raised against hnRNP K, hnRNP A2, and p54nrb (**F**). The spots of interest are indicated by arrows.

and the other (Coomassie blue) for preparative purposes. Reproducibility of the pattern was, thus, an essential point. To determine the best conditions for the precise location of spots of interest, we performed South-western blotting with ^{35}S-labelled proteins of HeLa nuclear extracts (see Materials and Methods). By exposing the blot on both sides with two phosphorimager screens we were able to obtain two radioactive images, one corresponding to the sum of impacts (^{32}P and ^{35}S), and one corresponding mainly to ^{32}P impacts (the ^{35}S contribution was eliminated by interposing an aluminium sheet). The specific spots revealed by South-western blotting could be localised by extrapolation of their positions in a coloured gel. Basically, the pattern of spots obtained with radioactivity ([^{35}S]methionine) or by Coomassie blue or silver staining was qualitatively identical (Fig. 4A and B), although their relative intensities could present variations. The reproducibility was confirmed over a large range of the total amount of proteins loaded. Proteins (40–500 µg) were loaded on the gel without loss of resolution. However, larger amounts resulted in lower resolution and increase of background signal. The quality of separation was optimal in the area corresponding to pH 4–8 and better in the extreme basic (pH 8–10) compared with extreme (pH 4–6) acidic region, and the density of spots was much higher in the acidic region.

Spots corresponding to radioactive signals as seen by South-western blotting with **T2** (Fig. 4C) were excised from Coomassie blue-stained gels (IPG pH 3–10), on which a maximum amount of 500 µg of nuclear HeLa extracts had been loaded. Despite the high reproducibility of the pattern from one gel to another, we could not exclude the risk of excising more than one protein by spot. The research software (see Materials and Methods) used to compare peptide mass data obtained from our sample to current bank data could provide the matching of two distinct proteins. However, the identity of the

Table 1. Recapitulative data from 2D-South-western blots obtained with HeLa extracts in the presence of various probes

Spot symbol		j	k	l	e	ab	i	y	w	z	x	
Relative abundance		7	27	8	12	25	3	8	7	1.5	1.5	S
Probe	T2	8.0	13.6	1.2	5.5	2.0	3.6	0.5	0.6	2.2	7.7	45.0
	T1(np)	4.8	8.8	0.3	2.8	0.3	2.6	0.8	0.8	0.4	2.9	24.0
	T1(op)	0.3	1.3	–	0.3	–	–	–	–	–	0.1	2.0
	T1(teg)	0.2	1.5	0.7	0.6	–	0.4	–	0.1	0.1	0.1	3.7
	T3′	1.5	3.0	1.3	0.5	5.3	0.6	4.1	0.8	0.5	2.1	19.7
	T1′(np)	8.3	16.4	1.9	4.1	2.9	0.5	3.9	0.2	0.2	3.5	42.0
	T1′(op)	9.7	7.8	1.0	1.6	3.2	0.8	1.7	0.1	1.1	10.8	37.8
	C2	1.6	4.6	0.1	1.2	0.3	–	0.1			6.4	14.3
	S2	1.7	8.2	–	2.4	0.6	0.8	0.1	0.4	0.3	2.7	17.2
	S1(np)	multiple spots										
	S3′	2.7	3.5	0.4	0.6	1.8	2.0	0.3	0.2	0.3	1.4	13.2
	S1′(np)	0.2	1.6	0.2	1.2	1.4	0.3	–	0.2	1.0	–	6.1
	S1′(op)	0.6	1.6	0.2	0.2	–	–	0.2	0.1	–	0.1	3.0

The signals (arbitrary units) of spots, calculated according to Materials and Methods, are reported in each row for a given probe. In the second row, a parameter indicating the relative abundance of the different proteins is obtained from ^{35}S exposition. In the last column, S stands for the sum of signals of all spots obtained with a given probe.

selected proteins was not aberrant in light of what could be expected from the DNA structures and on the basis of previously published work. Among eight spots excised, **k**, **l**, **e**, **ab**, **i**, **x**, **y**, **z1** (Fig. 4C), we clearly identified the first five spots (Table 2).

Members of the hnRNP family were identified as triplex-binding proteins

Five proteins belonging to the family of hnRNP (heteronuclear RiboNucleic Particles) were clearly identified. These were hnRNP K, L, E1, A2/B1 and I. The correspondence between the experimental and expected molecular weight and isoelectric point were within acceptable limits (18). Remarkably, these proteins were either acid (hnRNP K, J), neutral (hnRNP L, E) or basic (hnRNPs A2/B1, I). They are characterised by a variable number of nucleic acid-binding modules, such as KH (K homology) domains and RRM boxes, arginine- and glycine-rich sequences (19,20). Their sequence specificity as well as their relative affinities for RNA, single-stranded and duplex DNA, was shown to depend on the modular organisation of the protein (21). hnRNP K and I play a role in the pre-mRNA metabolism of transcripts containing cytidine- or pyrimidine-rich sequences (12). They are found as key components fulfilling various functions concerning RNA transportation and stabilisation, constitutive and alternative splicing and translational regulation. Their interactions with DNA allow them to play additional roles in regulation of transcription and replication.

Validation of protein identification by western blot analysis

Figure 4 shows the compared pattern of active spots obtained by South-western hybridisation with **T2** (Fig. 4C) and western hybridisation (Fig. 4F) performed on the same membrane upon 2D-electrophoresis. Antibodies (see Materials and Methods) specific for hnRNP K, A2/B1 and p54nrb were mixed for this experiment. The pattern was identical to the sum of spots revealed by individual western blots that had been previously performed with each specific antibody separately (not shown). The spots were marked with arrows and the identified corresponding protein was indicated. In these conditions, the western ECL signals showed an exact superimposition to the corresponding spots (**k**, **ab**, **ip**). In the case of spot **i**, it is interesting to note that a first identification attempt resulted in the erroneous finding of P54nrb (mol. wt 60 kDa, pI 9.5), a distinct abundant RNA-binding protein which migrated at a very close position in 3–10 IPG 2D-gels. In conditions of higher resolution, it turned out that the protein of interest migrated at a slightly lower position compared with p54nrb detected by western blot; it was unambiguously identified as hnRNP I by several independent experiments using MALDI-TOF, although we could not obtain confirmation of this data by western blot with anti-hnRNP I. This illustrates the risk of contamination of a protein of interest with another closely migrating protein of higher abundance in conditions of insufficient resolution.

Anti-hnRNP A2 revealed both hnRNP A2 (31 kDa) and B1 (38 kDa) which only differ by the insertion of a short sequence and are the products of alternative splicing. In the same way, hnRNP E1 (37 kDa, pI 7.07) and hnRNP E2 (38.6 kDa, pI 6.67) are products of alternative splicing from a common mRNA precursor and only differ by the insertion of nine amino acids. South-western blot performed with **T2** gave one spot common to the group of spots revealed by western blot with antibodies specific for hnRNP E (data not shown), matching better to hnRNP E1 according to the experimental pI value.

Table 2. Recapitulative data summarising mass spectrometry analysis

Research engine	Spot symbol	Identified protein	Number of matched peaks	Number of peaks	% Covered sequence	Exp mol. wt	Theo mol. wt	Exp pI	Theo pI
Profound	ab[1]	hnRNP A2/B1	13	16	31	35/40	36/37	9.5	8.97
expasy	″		12		36				
Msfit	″		8		24				
expasy	ab[2]		11	16	33				
Profound	i[1]	hnRNP I	9	11	20	55/60	57/59	9	9.2
expasy	″		8		16				
Mascot	″		9		/				
Profound	i[2]		9	11	19				
Profound	e[1]	hnRNP E1	5	27	17	40/45	37/38	7	6.66
expasy	″		8		34				
Profound	e[2]		5	12	17				
expasy	k	hnRNP K	15	33	37.5	60/70	51	5.2	5.4
Msfit	″		13		31				
Profound	1	hnRNP L	7	15	11	60/70	60	7	6.65
expasy	″		10	17	17				

Analysed proteins are designed by arabic letters related to the spots visualised in Figure 4. The following parameters are reported in the columns from left to right: Exp mol. wt and pI stand for experimental determinations of molecular weight and isoelectric point; Theo mol. wt and pI stand for theoretical data calculated from the amino acid sequence (www.expasy.org).
Indices 1 and 2 correspond to independent experiments.

Unidentified proteins

Some spots were subjected to repeated analysis by mass spectrometry but identification with pre-existing human proteins was not satisfactory (weak scores of percentages of recovered sequences or mostly unmatched peaks). Among those spots are triplex specific spots **w**, **y**, **z1** which are recognised when hybridised with different triplex sequences. Spot **j** was strongly suspected to correspond to a degradation product of hnRNP K according to published data (21). Spot **x**, which was observed with 55mer ODN containing a pyrimidine single strand, could not be identified. Low abundance is certainly a major obstacle but identification of hnRNP I was nevertheless possible from a gel where it was faintly visible when stained with Coomassie blue. Accurate localisation was also difficult in the case of spot **x** giving a poorly focused signal. These proteins probably correspond to as yet unknown proteins. Isoelectrofocalisation in basic range will be performed to try again to identify **w** and **z1**, **z2**.

DISCUSSION

This work was undertaken with the goal of designing a reliable method for screening DNA-binding proteins specific for unusual DNA structures. This involves the possibility to compare a panel of different nucleic probes, which differ from each other by a specific structural feature. This can be achieved by combining 2D-electrophoresis, South-western analysis and mass spectrometry. A typical small group of spots emerged with a set of probes sharing the ability to adopt intramolecular or intermolecular triplex structures. South-western analysis was already used in combination with 2D-electrophoresis and western blotting to confirm the recognition specificity of several DNA- and RNA-binding proteins studied in solution (21,22). Some pieces of evidence allowed us to validate the South-western method combined with 1D- or 2D-electrophoresis for the research of proteins binding to special DNA structures. The pattern of radiolabelled (hybridised) spots did not correspond to the most abundant proteins. Using a duplex probe containing the transcription factor sp1 recognition sequence (23), we obtained on 1D-gels a faint signal corresponding to the expected experimental molecular weight value of 110 kDa for sp1. In the same way, using poly(dC) as a probe, we revealed two bands corresponding to well known poly(C)-binding proteins: hnRNP (K and E).

Appropriate control probes were used to demonstrate that these interactions could not be ascribed to unfolded structures. (i) These spots were activated by various triplex probes (**l**, **k**, **e**, **i**, **w**, **y**, **z1** corresponding, respectively, to hnRNP L, K, E, I and three yet unidentified proteins); significant variations in intensity were observed as a function of the expected triplex stability, with decreasing signal intensities in the following order: CT(np) and CBrdU(op) > CT(op) and intramolecular > intermolecular triplexes. (ii) Some of the spots were also revealed with the corresponding third strand, albeit with lower intensities. While minor but reproducible variations were registered between triplexes in which only the sequence of the third strand differed, significant variations were observed when comparing their corresponding control as third strand alone. In the case of intermolecular triplexes, the profiles obtained with labelled **S1**(np) free or complexed with **D** were

very different. (iii) The South-western profile was maintained in the presence of a similar probe where intramolecular triplex has been covalently locked, using a psoralen cross-linked to the hairpin target (not shown).

Among the selected proteins hnRNP K and hnRNP E are not exclusively specific for triple-helices as other C-rich probes were also recognised: poly(dC) and human telomeric (C-rich) sequences (15). These were actually the two bands that were efficiently competed by an excess of poly(dC) in our experiments, at pH 7 [unstructured poly(dC)]. Interestingly, hnRNP I was highly labelled with triplexes in the pyrimidine motif, the associated signal was decreased with the unstable triplex control **C2**, while it was still present with the probe **T2** (teg). This was also the main protein revealed with intermolecular species, whatever the labelled strand.

We are inclined to conclude that the proteins detected so far actually accommodate several types of nucleic structures. The fact that some proteins (hnRNP J, K, L, I) were observed with both **T1** (TC) and **T3′** (TG-rich) argues in favour of triplex recognition rather than pyrimidine-rich single strand in the opened structure. Two proteins (spots **w** and **y**) were more specifically observed with G-rich triplexes, but could not be identified. They might correspond to still unknown proteins.

The five proteins identified as triplex-binding polypeptides are conserved along evolution and have a nucleic acid-binding domain KH or RRM. The selection of proteins belonging to the hnRNP family is not unexpected because of (i) their abundance among nuclear proteins and (ii) the intrinsic resistance of KH domain towards denaturation (19). Of special interest is hnRNP I which contains 4 RRM (24), and was highly labelled with triplex probes. Proteins identified in this study are likely good candidates for H-DNA recognition as they potentially bind to triplexes and single-stranded sequences to a lesser extent. Although the biological significance of these interactions is unclear, the selection of proteins involved in regulation of transcription [hnRNP K, (25–27)] or replication [hnRNP E2 (28)] actually makes H-DNA a potential player in the control of gene expression. The recognition mode of hnRNPs seems to use a variable register relative to nucleic acids sequence or structure. Their binding to various structures (D-loop, single-stranded and duplex DNA, possibly tetraplex such as G-quartets and triplexes) could constitute one of the mechanisms by which specific hnRNP could trigger and coordinate the regulation of various processes. Significantly, a yeast protein STM1 (mol. wt 30 kDa, pI 9.8), originally described to bind to G-quartet sequences, was recently hooked on the basis of its affinity for triplexes in the purine motif (29). Bomsztyk *et al*. (30) described hnRNP K as a molecular sensor interacting with RNA, DNA and proteins involved in transcriptional and translational gene regulation and signalling, in interdependent ways. Beyond the recognition of peculiar multi-stranded structures, the interactions of various ODNs with hnRNP proteins is of high interest with respect to the antisense and antigene strategies. These proteins might be essential partners influencing the intracellular localisation of ODN, by virtue of their tropism for the nuclear compartment (hnRNP L), or from their ability to shuttle between the nucleus and cytoplasm (hnRNP A, E, K, J, I among others) (31).

Garcia-Bassets *et al*. (32) recently purified four proteins of murine origin on the basis of their affinity for the pyrimidine strand potentially accessible (H-conformation) in d(GA)n.(TC)n sequences (but they do not bind to the purine strand, nor to the duplex). N-terminal peptide sequencing identified hnRNP K, L and I (also called PTB) and the fourth protein was not identified. Their affinity for three-stranded DNA has actually not been investigated in this report.

To validate our results, we performed preliminary experiments by Biosensor analysis with a recombinant protein containing the KH3 domain responsible for nucleic acid binding found in hnRNP (K and E) have shown that it actually interacts with triplex **T1** in the same incubation conditions as used for South-western analysis, validating the South-western method of selection of hnRNP (K and E) as triplex-binding proteins.

Despite the initial selection of proteins binding to triplex-forming sequences, we cannot be sure that the triplex structure is maintained in the nucleoprotein complex. We were unable to find a protein entirely specific for triplexes and totally devoid of affinity for another type of structure. Among less stable 55mers tested, some of them gave lower signals compared with triplex. Others gave similar signal as perfect triplexes, however, we could not exclude that an imperfect triplex responsible for the low temperature transition might be stabilised in the presence of such proteins; neither could be excluded the interaction of these proteins with the opened structure. Using the gel retardation approach, we picked up a protein specific for triplex form (5). However, a correlation was difficult to establish with information brought about by the South-western technique as we cannot ascertain that the protein seen by gel retardation was actually detectable by South-western blot.

Taking this data altogether we can conclude that selection of triplex-binding proteins results in the emergence of nucleic acid-binding factors characterised by a very large recognition spectrum. The identified proteins harbour different structural modules, which do not share at first sight a common feature. Future work will require comparative interaction studies using recombinant proteins containing different modules to understand the specific contribution of each structural feature involved in this recognition.

ACKNOWLEDGEMENTS

We thank L.Lacroix, J.L.Mergny for helpful discussion and providing antibodies and recombinant KH3 domain. We thank T.Garestier and C.Hélène for careful reading of the manuscript. We thank J.Nelson (Lynx therapeutics) for synthesising the 55mers triplex with np backbone for the third strand. We thank Pr. Celis, G.Dreyfuss, A.R.Krainer and A.Andino for the generous gift of antibodies. We are especially thankful to A.R.Krainer and M.Murray for welcoming F.G. in their laboratory. We thank V.Labas, J.S.Sun and C.Auvin for mass spectrometry opportunities and P.Alberti for Biosensor experiments. F.G. was a recipient of a MENRT fellowship.

REFERENCES

1. Mirkin,S.M. and Frank-Kamenetskii,M.D. (1994) H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 541–576.
2. Hampel,K.J., Ashley,C. and Lee,J.S. (1994) Kilobase-range communication between polypurine.polypyrimidine tracts in linear plasmids mediated by triplex formation: a braided knot between two linear duplexes. *Biochemistry*, **33**, 5674–5681.

3. Agazie,Y.M., Burkholder,G.D. and Lee,J.S. (1996) Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. *Biochem. J.*, **316**, 461–466.
4. Musso,M., Nelson,L.D. and Van Dyke,M.W. (1998) Characterization of purine-motif triplex DNA-binding proteins in HeLa extracts. *Biochemistry*, **37**, 3086–3095.
5. Guieysse,A.L., Praseuth,D. and Hélène,C. (1997) Identification of a triplex DNA-binding protein from human cells. *J. Mol. Biol.*, **267**, 289–298.
6. Kiyama,R. and Camerini-Otero,R.D. (1991) A triplex DNA-binding protein from human cells: purification and characterization. *Proc. Natl Acad. Sci. USA.*, **88**, 10450–10454.
7. Scheler,C., Lamer,S., Pan,Z., Li,X.P., Salnikow,J. and Jungblut,P. (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis*, **19**, 918–927.
8. Gevaert,K., Demol,H., Sklyarova,T., Vandekerckhove,J. and Houthaeve,T. (1998) A peptide concentration and purification method for protein characterization in the subpicomole range using matrix assisted laser desorption/ionization-postsource decay (MALDI-PSD) sequencing. *Electrophoresis*, **19**, 909–917.
9. Clauser,K.R., Hall,S.C., Smith,D.M., Webb,J.W., Andrews,L.E., Tran,H.M., Epstein,L.B. and Burlingame,A.L. (1995) Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. *Proc. Natl Acad. Sci. USA*, **92**, 5072–5076.
10. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
11. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
12. Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1992) Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.*, **12**, 164–171.
13. Kamma,H., Horiguchi,H., Wan,L., Matsui,M., Fujiwara,M., Fujimoto,M., Yazawa,T. and Dreyfuss,G. (1999) Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms. *Exp. Cell Res.*, **246**, 399–411.
14. Gamarnik,A.V. and Andino,R. (1997) Two functional complexes formed by KH domain containing proteins with the 5′ noncoding region of poliovirus RNA. *RNA*, **3**, 882–892.
15. Lacroix,L., Lienard,H., Labourier,E., Djavaheri-Mergny,M., Lacoste,J., Leffers,H., Tazi,J., Hélène,C. and 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.
16. Shevchenko,A., Wilm,M., Vorm,O. and Mann,M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.
17. Escude,C., Giovannangeli,C., Sun,J.S., Lloyd,D.C., Chen,J.K., Gryaznov,S.M., Garestier,T. and Hélène,C. (1996) Stable triple helices formed by oligonucleotide N3′→P5′ phosphoramidates inhibit transcription elongation. *Proc. Natl Acad. Sci. USA*, **93**, 4365–4369.
18. Perrot,M., Sagliocco,F., Mini,T., Monribot,C., Schneider,U., Shevchenko,A., Mann,M., Jeno,P. and Boucherie,H. (1999) Two-dimensional gel protein database of Saccharomyces cerevisiae (update 1999). *Electrophoresis*, **20**, 2280–2298.
19. Dejgaard,K. and Leffers,H. (1996) Characterisation of the nucleic-acid-binding activity of KH domains. Different properties of different domains. *Eur. J. Biochem.*, **241**, 425–431.
20. Leffers,H., Dejgaard,K. and Celis,J.E. (1995) Characterisation of two major cellular poly(rC)-binding human proteins, each containing three K-homologous (KH) domains. *Eur. J. Biochem.*, **230**, 447–453.
21. Dejgaard,K., Leffers,H., Rasmussen,H.H., Madsen,P., Kruse,T.A., Gesser,B., Nielsen,H. and Celis,J.E. (1994) Identification, molecular cloning, expression and chromosome mapping of a family of transformation upregulated hnRNP-K proteins derived by alternative splicing. *J. Mol. Biol.*, **236**, 33–48.
22. Won,K.A. and Baumann,H. (1991) NF-AB, a liver-specific and cytokine-inducible nuclear factor that interacts with the interleukin-1 response element of the rat alpha 1-acid glycoprotein gene. *Mol. Cell. Biol.*, **11**, 3001–3008.
23. Shou,Y., Baron,S. and Poncz,M. (1998) An Sp1-binding silencer element is a critical negative regulator of the megakaryocyte-specific alphaIIb gene. *J. Biol. Chem.*, **273**, 5716–5726.
24. Perez,I., McAfee,J.G. and Patton,J.G. (1997) Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry*, **36**, 11881–11890.
25. Krecic,A.M. and Swanson,M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.*, **11**, 363–371.
26. Miau,L.H., Chang,C.J., Shen,B.J., Tsai,W.H. and Lee,S.C. (1998) Identification of heterogeneous nuclear ribonucleoprotein K (hnRNP K) as a repressor of C/EBPbeta-mediated gene activation. *J. Biol. Chem.*, **273**, 10784–10791.
27. Tomonaga,T. and Levens,D. (1996) Activating transcription from single stranded DNA. *Proc. Natl Acad. Sci. USA*, **93**, 5830–5835.
28. Parsley,T.B., Towner,J.S., Blyn,L.B., Ehrenfeld,E. and Semler,B.L. (1997) Poly (rC) binding protein 2 forms a ternary complex with the 5′-terminal sequences of poliovirus RNA and the viral 3CD proteinase. *RNA*, **3**, 1124–1134.
29. Nelson,L.D., Musso,M. and Van Dyke,M.W. (2000) The yeast STM1 gene encodes a purine motif triple helical DNA-binding protein. *J. Biol. Chem.*, **275**, 5573–5581.
30. Bomsztyk,K., Van Seuningen,I., Suzuki,H., Denisenko,O. and Ostrowski,J. (1997) Diverse molecular interactions of the hnRNP K protein. *FEBS Lett.*, **403**, 113–115.
31. Michael,W.M., Eder,P.S. and Dreyfuss,G. (1997) The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J.*, **16**, 3587–3598.
32. Garcia-Bassets,I., Ortiz-Lombardia,M., Pagans,S., Romero,A., Canals,F., Avilis,F.X. and Azorin,F. (1999) The identification of nuclear proteins that bind the homopyrimidine strand of d(GA.TC)n sequences, but not the homopurine strand. *Nucleic Acids Res.*, **27**, 3267–3275.