

AND-1, a natural chimeric DNA-binding protein, combines an HMG-box with regulatory WD-repeats

Andreas Köhler, Marion S. Schmidt-Zachmann and Werner W. Franke*

Division of Cell Biology, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

*Author for correspondence

SUMMARY

Using a specific monoclonal antibody (mAb AND-1/23-5-14) we have identified, cDNA-cloned and characterized a novel DNA-binding protein of the clawed toad, *Xenopus laevis*, that is accumulated in the nucleoplasm of oocytes and various other cells. This protein comprises 1,127 amino acids, with a total molecular mass of 125 kDa and a pI of 5.27. It is encoded by a mRNA of ~4 kb and contains, in addition to clusters of acidic amino acids, two hallmark motifs: the amino-terminal part harbours seven consecutive 'WD-repeats', which are sequence motifs of about 40 amino acids that are characteristic of a large group of regulatory proteins involved in diverse cellular functions, while the carboxy terminal portion possesses a 63-amino-acid-long 'HMG-box', which is typical of a family of DNA-binding proteins involved in regulation of chromatin assembly, transcription and replication. The DNA-binding

capability of the protein was demonstrated by DNA affinity chromatography and electrophoretic mobility shift assays using four-way junction DNA. Protein AND-1 (acidic nucleoplasmic DNA-binding protein) appears as an oligomer, probably a homodimer, and has been localized throughout the entire interchromatinic space of the interphase nucleoplasm, whereas during mitosis it is transiently dispersed over the cytoplasm. We also identified a closely related, perhaps orthologous protein in mammals. The unique features of protein AND-1, which is a 'natural chimera' combining properties of the WD-repeat and the HMG-box families of proteins, are discussed in relation to its possible nuclear functions.

Key words: Chromatin, DNA-binding protein, HMG-box, Nucleoplasm, WD-repeat

INTRODUCTION

The nucleoplasm is a large subcellular compartment containing a pool of soluble proteins, protein complexes and particles that can be rapidly recruited, in a regulated way, to serve specific nuclear functions such as chromatin formation, replication, transcription, ribonucleoprotein assembly, polyadenylation and splicing. With increasing availability of amino acid sequence information about such nucleoplasmic proteins, it has become clear that they can be grouped according to certain functional roles or functional sequence motifs. Examples of important groups of the latter category of nuclear proteins include those characterized by zinc-binding domains (reviewed by Harrison, 1991; Saurin et al., 1996), leucine zipper (reviewed by Busch and Sassone-Corsi, 1990; Harrison, 1991), WD-repeats (reviewed by van der Voorn and Ploegh, 1992; Neer et al., 1994; Neer and Smith, 1996) and HMG-boxes (reviewed by Johns, 1982; Landsman and Bustin, 1993; Laudet et al., 1993).

In the present study we report the first example of a nucleoplasmic protein (M_r 125,000) that is acidic and combines hallmark features of two different regulatory protein families, the DNA-binding HMG-box and the WD-repeat protein families. We show that this protein, which can bind intensely to DNA, is a widespread, probably ubiquitous

nucleoplasmic constituent, and we discuss the possibility that its domain dualism reflects dual or even multiple regulatory interactions.

MATERIALS AND METHODS

Biological material

Clawed toads (*Xenopus laevis*) were purchased from the South African Snake Farm (Fish Hoek, South Africa). Cultured cells derived from *Xenopus laevis* kidney epithelium (XLKE, line A6; American Type Culture Collection, ATCC, CCL 102) were grown at 27°C in 78.3% Dulbecco's minimal essential medium (DMEM; ICN Biomedicals, Costa Mesa, CA, USA), supplemented with 8.7% FCS and 13% H₂O (e.g. 90 ml DMEM + 10 ml FCS + 15 ml H₂O). Cultured cells of the human hepatocellular carcinoma line PLC (ATCC, CRL 8024) were grown at 37°C in 90% DMEM supplemented with 10% FCS.

Isolation and fractionation of oocyte nuclei from *Xenopus laevis*

Nuclei of mature (stages IV-VI; Dumont, 1972) *Xenopus laevis* oocytes were obtained by mass isolation, a large-scale procedure described by Scalenghe et al. (1978) and modified by Kleinschmidt and Franke (1982). The subsequent fractionation of the intact nuclei was performed as described by Hügle et al. (1985), resulting in a 'low speed pellet' (LSP; 3,500 g, 15 minutes) containing nucleoli, nuclear envelopes and chromosomes, a 'high speed pellet' (HSP; 100,000 g,

1 hour), enriched in preribosomal particles and a 'high speed supernatant' (HSS), containing soluble nuclear proteins.

Generation of monoclonal antibodies

Monoclonal antibodies were raised essentially according to the method of Köhler and Milstein (1975). A 7-week-old female Balb/c mouse was immunized with 100 µg HSP-material from *Xenopus laevis* oocyte nuclei, prepared as described above. After two booster injections, the spleen cells were harvested and fused with cells of the mouse myeloma line, P3X63-Ag8.653, at a ratio of 3:1 in the presence of 40% PEG 4,000. Antibody-producing hybridoma cell lines were selected essentially as described by Schmidt-Zachmann et al. (1984). Immunoglobulin subclasses were determined by ELISA with subclass-specific antibodies (Medac, Hamburg, Germany) and peroxidase-coupled secondary antibodies (Sigma, Munich, Germany).

One of the mAbs obtained in this preparation was named mAb AND-1/23-5-14. This antibody of the IgG1-subclass was used for the experiments described in the present study. The hybridoma cell line 23-5-14 was also propagated as ascites in Balb/c mice.

Preparation of XLKE-A6 cell extracts

Extracts from XLKE-A6 cells were prepared by incubating the cells of one confluent 10 cm cell culture dish (2×10^7 cells) in 1 ml lysis buffer (100 mM Na-MES pH 6.5, 250 mM NaCl, 0.5 mM MgCl₂, 5 mM EDTA, 0.5% Triton X-100, 0.02% NaN₃, 50 µM Pefa-Block, 20 µM Pepstatin, 20 µM Leupeptin) for 2 minutes at room temperature. The lysed cells were cleared from cellular debris by centrifugation (14,000 g, 10 minutes, 4°C).

Gel electrophoresis and immunoblotting

Protein fractions were separated by SDS-PAGE according to Thomas and Kornberg (1975). The two-dimensional separation of proteins was carried out by isoelectric focussing (IEF; O'Farrell, 1975) in the first dimension and by SDS-PAGE in the second dimension. The polypeptides were transferred to nitrocellulose and visualized by Poinceau S staining. The nitrocellulose membranes were blocked in TBS containing 0.05% Tween (TBST) and 5% non-fat dry milk for 1 hour and then incubated at room temperature with culture supernatant, diluted 1:2 with TBST for 1 hour. After washing with TBST, bound antibodies were either detected by colour reaction after incubation with alkaline-phosphatase-coupled secondary anti-mouse IgG1 antibodies (Dianova, Hamburg, Germany) or by chemiluminescence using the ECL-system (Amersham, Braunschweig, Germany) with horse radish peroxidase (HRP)-coupled secondary anti-mouse IgG1 antibodies (Dianova). In the case of ECL detection, the incubation with the HRP-coupled antibodies was carried out in the presence of 5% non-fat dry milk in TBST.

Immunofluorescence microscopy

For immunofluorescence microscopy studies on cultured cells, the cells grown on coverslips were either fixed with methanol (5 minutes, -20°C) and acetone (30 seconds, -20°C) or with 2% formaldehyde, freshly prepared from paraformaldehyde (in PBS, pH 7.4) for 20 minutes. Following fixation, cells were washed twice with PBS and incubated with mAb AND-1/23-5-14 (ascites fluid, 1:200 diluted in PBS) for 30 minutes. After several PBS washes, the cells were incubated for 15 minutes with Texas Red-labeled goat anti-mouse Ig secondary antibodies (Dianova) diluted 1:100 in PBS. The cells were then washed again with PBS, dehydrated in ethanol, air-dried and mounted with Elvanol (Hoechst, Frankfurt, Germany).

Cryosections (5 µm) through *Xenopus laevis* ovary, epidermis and other tissues were fixed with 2% formaldehyde. Incubation times of primary and secondary antibodies were as described above. Micrographs were taken with an Axiophot microscope (Zeiss, Oberkochen, Germany).

Density gradient analysis and gel filtration

Analysis of the HSS-fraction by density gradient centrifugation was

done essentially as described before (Hügler et al., 1985) with minor modifications. 600 µl of the HSS-fraction were fractionated across a 5%-30% linear sucrose gradient, made in '5:1'-medium (10 mM Tris-HCl pH 7.4, 83 mM KCl, 17 mM NaCl, 2 mM MgCl₂, 2.5 mM DTT, 50 µM Pefa-Block). 15 fractions of 0.8 ml each were collected from the top (light) to the bottom (heavy) of the gradient. Marker proteins (BSA, catalase, thyroglobulin; all from Pharmacia, Freiburg, Germany) were separated on parallel gradients.

For gel filtration analysis the lysis supernatant derived from one 10 cm culture dish of confluent grown XLKE-A6 cells was applied on a Superose 12 column (HR 10-30; Pharmacia) at room temperature. For calibration, reference proteins (thyroglobulin, ferritin, catalase, aldolase and BSA; all from Pharmacia) were fractionated together with cellular extracts. Proteins were eluted (48 fractions, 0.2 ml each) in 100 mM Na-MES, pH 6.5, 250 mM NaCl, 0.02% NaN₃, 0.5 mM MgCl₂, 5 mM EDTA. After methanol precipitation, 1/6 of each fraction was analyzed by SDS-PAGE and immunoblotting.

Immunoprecipitation of protein AND-1

Immunoprecipitations were performed using whole, homogenized *Xenopus* oocyte nuclei, HSS-material, XLKE-A6 cell lysates and total egg extracts. The latter ones were prepared as described previously by Cordes et al. (1993). Antibodies were first coupled by pre-incubation (1 hour, 4°C) of mAb AND-1/23-5-14 ascites fluid (50 µl, diluted with 450 µl PBS) with 50 µl pre-swollen protein-G-Sepharose (Pharmacia). Protein solutions were also pre-incubated with protein-G-Sepharose (1 hour, 4°C) to avoid unspecific binding during the immunoprecipitation process. During end-over-end rotation (2 hours, 4°C), the pre-cleared protein sample was incubated with the mAb AND-1/23-5-14-Protein G complex. After low-speed centrifugation (800 g, 5 minutes) the resulting supernatant was precipitated with acetone and prepared for SDS-PAGE. The Sepharose beads with the bound immune complexes were washed five times with PBS, then once each with PBS containing 0.1% Triton X-100 and with PBS, and finally boiled in SDS-sample buffer. The solubilized proteins from the precipitate were analyzed by SDS-PAGE, together with the starting material and the reserved supernatant fraction.

RNA-isolation and northern blot hybridization

Total RNA from *Xenopus* ovaries was prepared as described by Chomczynski and Sacchi (1987). Poly(A)⁺ RNA was prepared using the mRNA Purification Kit (Pharmacia). For northern blot analysis RNA was separated on a 1% agarose-gel containing 0.6 M formaldehyde, transferred to Biotrans A filters (Pall, Dreieich, Germany), hybridized, washed and visualized essentially as described by Heid et al. (1994). For hybridization, RNA blots were probed with a 1.59 kb ³²P-labeled antisense cRNA derived from *EcoRI* digestion of pBS AND-1-*Xen* and prepared using RNA-polymerase T7.

Isolation and characterization of AND-1-cDNA clones

A λ unizap cDNA expression library from total *Xenopus laevis* ovary (Stratagene, Heidelberg) was screened using mAb AND-1/23-5-14. Positive plaques were identified by ECL-detection (Amersham). Two positive clones were selected, plaque-purified and released from the phages by *in vivo* excision, according to the manufacturer's protocol. The two resulting cDNAs cloned in pBluescript were investigated by restriction mapping and preliminary sequence analysis, revealing that the ~2.6 kb cDNA was identical to the 3'-end of the ~4 kb cDNA. The entire ~4 kb cDNA sequence of pBS-AND-1-*Xen* was analyzed by constructing a set of deletion clones using the Double-stranded nested deletion kit (Pharmacia) and subsequent sequencing. The sequence analysis was performed using the HUSAR (Heidelberg Unix Sequence Analysis Resources) software program package.

Transfection and *in vitro* transcription/translation

For transient expression in human PLC cells, the AND-1 cDNA was subcloned as a *Apal/NotI*-fragment into the mammalian expression

vector pRc/CMV (Invitrogen via ITC Biotechnology, Heidelberg, Germany). Transfections, using 10 µg of DNA per 3.5 cm tissue culture dish, were carried out according to the method described by Chen and Okayama, (1987). Transfected cells were analyzed by immunofluorescence microscopy 24 hours after removal of the DNA-Ca²⁺ precipitate.

The same construct was used for the production of [³⁵S]methionine-labeled protein AND-1 in vitro using the TNT™ Coupled Reticulocyte Lysate System (Promega via Boehringer Ingelheim Bioproductions, Heidelberg, Germany).

Identification of AND-1 transcripts by PCR

AND-1 transcripts in different tissues and cell lines from *Xenopus* and rat were detected by PCR. Two oligonucleotides (AND-1 sense 1, corresponding to nucleotides 1,835-1,857; AND-1 antisense 1, reverse complementary to nucleotides 2,212-2,235) were designed corresponding to high homologous regions within the AND-1 cDNA and an EST sequence from rat adrenal chromaffin PC-12 cells (EST 108147; Lee et al., 1995). RT-cDNA from PC-12 cells and from rat brain and liver were kindly provided by Dr R. Leube (Johannes Gutenberg University, Institute of Anatomy, Mainz, Germany; for preparation of RT-cDNA see Leube et al., 1994). Suitable templates for PCR were also obtained by preparation of phage DNA from λ cDNA expression libraries of *Xenopus laevis* kidney, liver and ovary (Stratagene) or human fetal brain and epidermis (Clontech, Heidelberg, Germany). The amplification was performed under the following conditions: 5 µl of template was added to the reaction, containing 10 pmol of each primer and 0.2 mM dNTP in 1× reaction buffer. 1 unit Taq-polymerase (Boehringer Mannheim, Germany) was added to a final volume of 50 µl. The following PCR-profile was used: 1× 94°C/3 minutes; 15× (92°C/30 seconds, 55°C/1 minute (-0.5°C/cycle), 72°C/1 minute); 25× (92°C/30 seconds, 50°C/1 minute, 72°C/1 minute (+2 seconds/cycle)); 1× 72°C/5 minutes. The reactions were performed using a Minicycler (MJ research, via Biozym Diagnostics, Hess. Oldendorf, Germany).

DNA cellulose chromatography

The DNA cellulose chromatography was performed essentially as described by Miller et al. (1989). A column was packed with 1 ml cellulose, coupled with native, double-stranded calf thymus DNA (Pharmacia) and pre-swollen in buffer A (15 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM EDTA, 1 mM DTT, 15% glycerol, 0.05% NP-40). 2 ml of the HSS-fraction, diluted with 8 ml buffer A, were loaded. The column was washed with 10 ml of loading buffer A and bound proteins were subsequently eluted with successive applications of 2 ml of buffer A containing 100 mM, 150 mM, 200 mM, 250 mM and 1.5 M NaCl, respectively. Fractions of 0.5 ml were collected, proteins were precipitated with methanol and subsequently analyzed by SDS-PAGE.

Recombinant expression of the AND-1 HMG-box

The AND-1-HMG-box was expressed and purified from bacteria using the Qiaexpress system (Qiagen, Hilden, Germany). The corresponding cDNA was subcloned as a 296 nt *Pst*I-fragment (nt 3,099-3,395) in the pQE 30 vector. Following high expression in *E. coli* strain M15[pREP4], the protein, containing a 6× His affinity tag, was purified on Ni-NTA resin under denaturing conditions. The protein was subsequently dialyzed step by step in buffer BC-0 (20 mM Tris-HCl, pH 7.9, 20% glycerol, 0.2 mM EDTA, 5 mM MgCl₂) containing 6 M, 4 M, 2 M, 1 M and 0 M urea, respectively.

Electrophoretic mobility shift assay with four-way junction DNA

The preparation of the four-way junction DNA and the electrophoretic mobility shift assay was done as described by Bianchi (1988) and Kuhn et al. (1994). 400 and 600 ng of the bacterial-expressed AND-1 HMG-box was used in the assay. 20 ng of the full-length bac-

culovirus-expressed mouse UBF1 (kindly provided by Dr Renate Voit, this center, Division of Molecular Biology of the Cell II) containing five HMG-boxes was used as a control.

RESULTS

Identification of a novel nuclear protein in *Xenopus laevis* oocytes and somatic cells

In the course of our studies on proteins stored in the nucleus ('germinal vesicle') of *Xenopus laevis* oocytes, such as the histone-binding proteins nucleoplasmin (Krohne and Franke, 1980a,b), and N1/N2 (Kleinschmidt and Franke, 1982; Kleinschmidt et al., 1985, 1986), high mobility group protein HMG-A (Kleinschmidt et al., 1983a), a nuclear actin-binding protein (Ankenbauer et al., 1989), the Mg²⁺-ATPase and the proteasome particles (Kleinschmidt et al., 1983b, Hügler et al., 1983; Peters et al., 1990, 1994), we had come across a protein with several unusual properties. One of the murine monoclonal antibodies, mAb AND-1/23-5-14 (IgG1), which was obtained after immunization with high speed pellet (HSP) proteins from isolated oocyte nuclei, reacted in immunoblotting experiments specifically with a polypeptide of *M_r* ~125,000 identifiable in whole oocyte nuclei (Fig. 1, lane 1), in the HSP fraction containing particulate nuclear material (Fig. 1, lane 2), in the high speed supernatant (HSS) fraction containing the soluble nuclear proteins (Fig. 1, lane 3) and in total egg extracts (Fig. 1, lane 4). From these experiments we gained the impression that in oocyte nuclei the protein was present in a particulate-bound and in a soluble form. To examine the occurrence of this protein in somatic cells, cultured *Xenopus laevis* kidney epithelial cells (XLKE-A6) were lysed in a buffer containing 0.5% Triton X-100 and 250 mM NaCl, and solubilized proteins were fractionated by centrifugation. Using immunoblotting on SDS-PAGE-separated proteins, the same polypeptide of *M_r* 125,000 was detected and completely recovered among the soluble proteins (Fig. 1, lanes 5, 6). Due to the minor occurrence of the protein in XLKE-A6 cells (compared to *Xenopus* oocytes and eggs) lane 5 had to be overloaded. To allow a comparison between the novel protein and other well-characterized nuclear *Xenopus* proteins, the immunoblot shown in Fig. 1A' was reprobed with antibodies recognizing the histone-binding protein N1/N2 (kindly provided by Dr J. Kleinschmidt, this center, Division of Applied Tumor Virology; see Kleinschmidt et al., 1985, 1986). The result is shown in Fig. 1A''.

After separation of HSS-material by two-dimensional gel electrophoresis it was also possible to compare the immunoreactive protein with protein N1/N2 or nucleoplasmin on Coomassie blue-stained gels (Fig. 1B). It appeared as a faint, but Coomassie-visible spot (marked by an arrow), whereas N1/N2 and nucleoplasmin (Np) were the most abundant proteins of this fraction. Immunoblot analysis of the separated proteins showed that the polypeptide recognized by mAb AND-1/23-5-14 was an acidic protein with a pI of ~5.5 (arrow, Fig. 1B').

Sections through frozen *Xenopus* oocytes were examined by immunofluorescence microscopy with mAb AND-1/23-5-14, and the antigenic protein was detected only in the nucleus (Fig. 2A,A'). The protein was also localized in nuclei of certain other tissues such as epidermis (Fig. 2B,B') and in nuclei of monolayer cultures of XLKE-A6 cells (Fig. 2C-G), where it appeared in

Fig. 1. Identification of the polypeptide recognized by mAb AND-1/23-5-14 in different fractions from *Xenopus* oocyte nuclei, eggs and somatic cells.

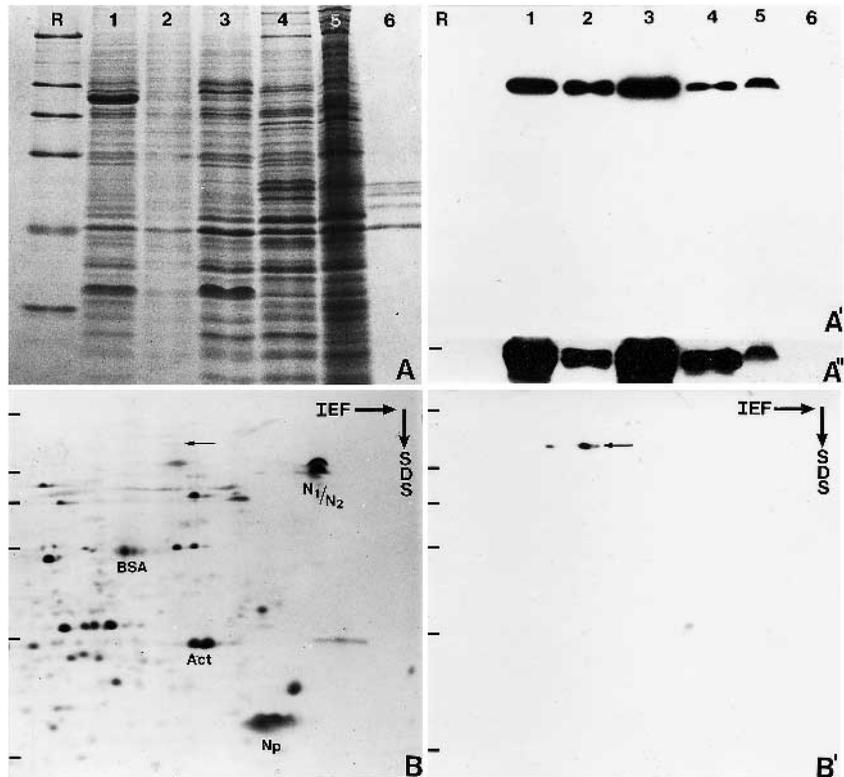
(A) Coomassie blue staining of the SDS-PAGE-separated polypeptides. (A') Corresponding immunoblot using mAb AND-1/23-5-14. Bound antibodies were detected by ECL (Amersham).

(R) Reference proteins (from top to bottom in kDa): 205, 116, 97, 66, 45 and 29. Lane 1, proteins from whole isolated oocyte nuclei. Lane 2, proteins of high speed pellet (HSP) fraction from oocyte nuclei (1/75 of the total HSP-fraction). Lane 3, proteins of the corresponding high speed supernatant (HSS) fraction from oocyte nuclei (1/75 of the total HSS fraction).

Lane 4, proteins from the 100,000 g supernatant fraction of non-activated *Xenopus* eggs ('egg extract'). Lane 5, supernatant fraction containing the soluble proteins from cultured *Xenopus* kidney epithelial cells of line A6 (XLKE-A6) lysed in a buffer containing 0.5% Triton X-100 and 250 mM NaCl and subsequently fractionated by centrifugation. Lane 6, corresponding pellet fraction.

The monoclonal antibody mAb AND-1/23-5-14 reacted selectively with a polypeptide of $M_r \sim 125,000$, which is enriched in *Xenopus* oocytes and eggs. The protein was also detectable in the somatic XLKE-A6 cells (lane 5) in a form that was completely solubilized in the presence of 250 mM NaCl. (A'') For comparison, the immunoblot shown in A' was reprobbed with antibodies recognizing protein N₁/N₂ (Kleinschmidt et al., 1986).

(B) The soluble proteins from *Xenopus* oocyte nuclei present in the HSS-fraction were analyzed by two-dimensional gel electrophoresis (iso-electric focussing/SDS-PAGE) and subsequently stained with Coomassie Blue. (B') Corresponding immunoblot using mAb AND-1/23-5-14. The reactive polypeptide has an apparent pI of ~ 5.5 and is denoted by a small arrow. The smaller dot in the immunoblot (B') may represent a more basic variant. The polypeptide recognized by mAb AND-1/23-5-14 is visible as a faint spot after staining with Coomassie blue. For comparison, the abundant soluble proteins N₁/N₂ and Nucleoplasmin (Np) are indicated in B. Reference proteins used for co-electrophoresis were α -actin (Act) and BSA. Horizontal bars on the left indicate the position of reference proteins (same as in A).



interphase cells with a dispersed, sometimes finely granular distribution throughout the nucleoplasm. During mitosis (prometaphase to anaphase, Fig. 2D-G) the immunostained protein was dispersed over the cytoplasm, showing in metaphase and anaphase a distribution that somewhat accentuated the surfaces of the condensed chromosomes (Fig. 2E,F). In telophase the protein rapidly re-accumulated within the forming daughter nuclei (Fig. 5G). The protein resembled certain sequence-specific transcription factors in some of its properties and distribution during mitosis (cf. Martinez-Balbás et al., 1995).

Because of its major properties we refer to the protein by the acronym AND-1 (acidic nucleoplasmic DNA-binding protein).

The native state of soluble protein AND-1

The state of soluble protein AND-1 was analyzed by density gradient centrifugation. The protein could be identified in a complex with an S value of ~ 7.5 (Fig. 3A,A'), which when analyzed by gel filtration, appeared with an apparent molecular mass (M_{app}) of 800,000 (Fig. 3B). According to Siegel and Monty (1966) this allows an estimate for the relative molecular mass (M_r) of $\sim 240,000$ for the soluble form.

The complex containing protein AND-1 could also be immunoprecipitated from protein extracts of whole *Xenopus* oocyte nuclei, the HSS fraction of soluble nuclear proteins, egg extracts and XLKE-A6 cell lysates, but no co-precipitating proteins were detected in near stoichiometric amounts (Fig. 4).

From these data, we conclude that the bulk of the soluble protein AND-1 appears as a homo-oligomer, probably a homodimer.

Isolation and analysis of cDNA

To isolate the cDNA encoding protein AND-1, a λ cDNA expression library of *Xenopus laevis* ovary was screened with mAb AND-1/23-5-14. We obtained two clones with inserts of ~ 2.7 kb and ~ 4 kb. The smaller clone was found to be identical to the 3'-end of the larger cDNA clone. The latter was denoted pBS-AND-1-*Xen* and further analyzed.

Total RNA and poly(A)⁺ RNA from *Xenopus* ovary tissue were probed in northern blot experiments with a 1.59 kb antisense cRNA derived from pBS-AND-1-*Xen*. We obtained a strong signal corresponding to a mRNA of ~ 4 kb (Fig. 5A), indicating that we had isolated a cDNA clone of full or nearly full length.

The completeness of the cDNA clone was demonstrated in two types of expression experiments. First, the ~ 4 kb cDNA was subcloned into the mammalian expression vector pRc/CMV and subjected to coupled transcription and translation in vitro. After separation of the [³⁵S]methionine-labeled translation products by one- (Fig. 5B) or two-dimensional gel electrophoresis (Fig. 5C) the major product identified showed SDS-PAGE mobility and an apparent pI that were practically identical to those of the native protein AND-1 (compare Fig. 5B,C with Fig. 1). This was confirmed by co-electrophoresis.

After separation of the radiolabeled translation product by two-dimensional gel electrophoresis, the native protein AND-1, detected by alkaline phosphatase reaction, and the product of *in vitro* translation, detected by autoradiography, were found to co-localize in the same polypeptide spot (results not shown).

In transiently transfected cultured human PLC cells the protein encoded by the cDNA clone showed the same nucleoplasmic localization as the native protein AND-1 (see Fig. 5D,D').

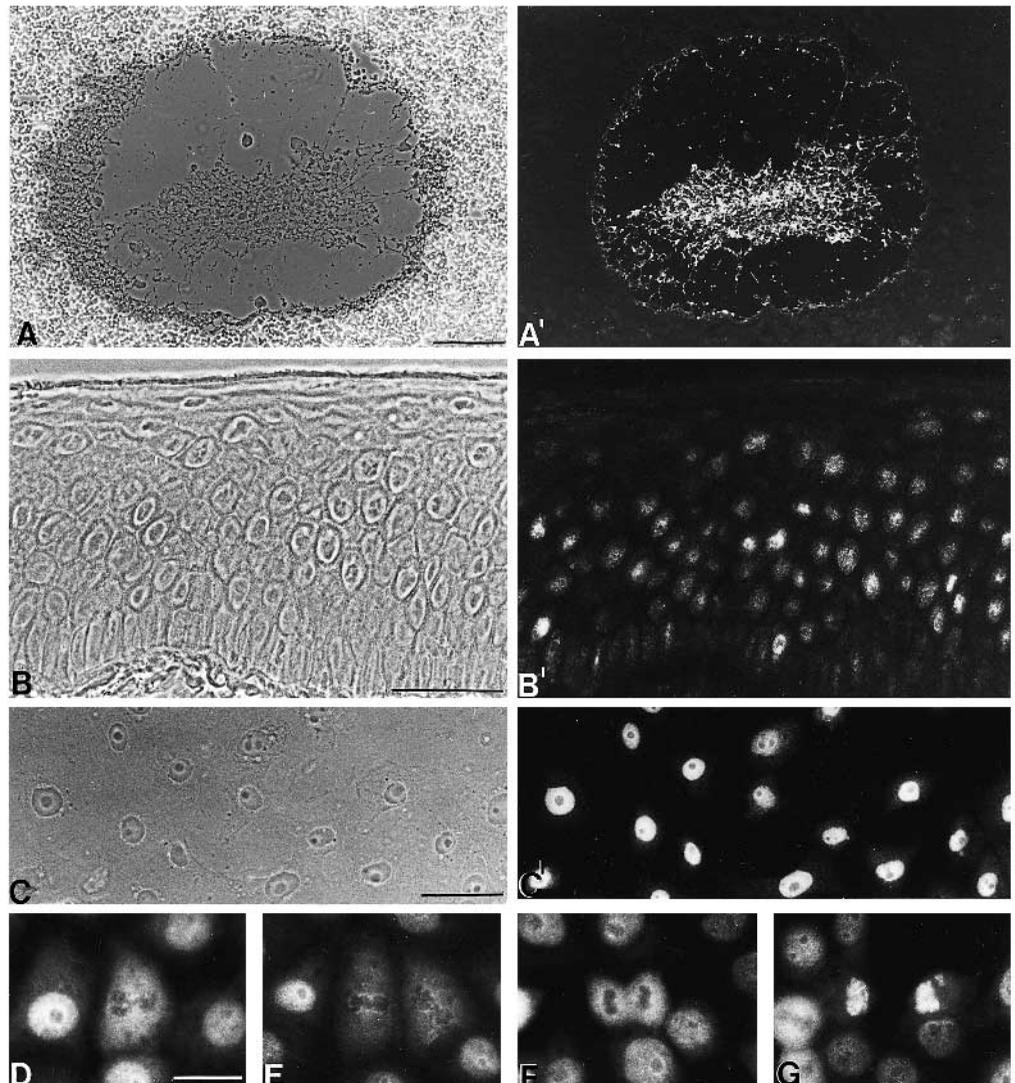
The nucleotide sequence of the cDNA was determined in both directions (for data, see EMBL data base, accession no. X98884), and the amino acid (aa) sequence deduced from it is shown in Fig. 6A. The clone of 3,978 nucleotides, in agreement with the ~4 kb size of the mRNA as detected by northern blot analysis (cf. Fig. 5A), contained a 5'-untranslated region of 81 nucleotides, with an in-frame stop codon (nt 17-19) upstream of the first possible start codon (nt 82-84), an open reading frame of 3,381 nt, and a 3' region of 515 untranslated nucleotides with a polyadenylation signal (AATAAA) 12 nt before the poly(A) tail, of which 22 nt were included in the clone. The open reading frame encoded a polypeptide of 1,127 amino acids with a calculated molecular mass of 124.52 kDa

and a pI of 5.27, both in close agreement with the data found for the native protein and indicating that the mature polypeptide has not been considerably altered by modifications that would have resulted in changes of its electrophoretic properties.

Protein AND-1 is a member of both the WD-repeat and the HMG-box families

Data base searches revealed that the protein was novel, but also that its amino acid sequence contained some striking features (Figs 6, 7): the amino-terminal segment consists of seven WD repeats, a type of repeating unit with a basic number of ~40 aa that often ends with a tryptophan (W) and an aspartic acid (D) residue (for references see van der Voorn and Ploegh, 1992; Neer et al., 1994). Members of the WD repeat protein family have been shown to be involved in the regulation of diverse fundamental cell functions, including nuclear ones (see Discussion). Of particular interest is the relationship of protein AND-1 to the SepB protein of *Aspergillus nidulans*, which is known to be involved in chromosome segregation and cytokinesis initiation (Harris and Hamer, 1995). An alignment of the seven WD repeats found in protein AND-1 compared to the

Fig. 2. Immunolocalization of the protein recognized by mAb AND-1/23-5-14 in *Xenopus* tissues and cultured XLKE-A6 cells. (A) Phase-contrast micrograph of a frozen section of *Xenopus* ovary. (A') Corresponding epifluorescence photograph. (B) Phase-contrast micrograph of a frozen section of *Xenopus* epidermis. (B') Corresponding epifluorescence photograph. (C) Phase-contrast optics of formaldehyde-fixed XLKE-A6 cells. (C') Same field in epifluorescence optics. The protein was always found exclusively in the nucleoplasm of the cells. The reason for the reaction heterogeneity from cell to cell is not clear. (D-G) Epifluorescence photographs showing the distribution of the protein recognized by mAb AND-1/23-5-14 during mitosis of XLKE-A6 cells. (D) Prometaphase. (E) Metaphase. (F) Anaphase. (G) Telophase. Note the distribution of the protein during mitosis, especially its displacement from mitotic chromatin (cf. E and F), and its rapid re-accumulation in the daughter nuclei. Bars: 50 μ m (A-C'); 20 μ m (D-G).



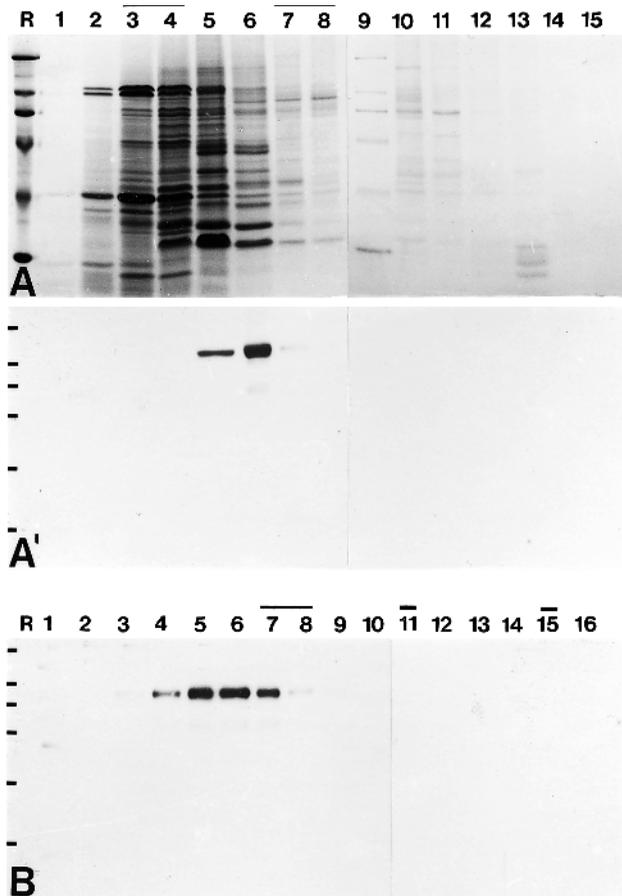


Fig. 3. Analysis of the native state of soluble protein AND-1 by sucrose gradient centrifugation and gel filtration. (A) Soluble proteins from *Xenopus* oocyte nuclei (HSS-fraction) were fractionated after sucrose gradient centrifugation, separated by SDS-PAGE and stained with Coomassie blue. (A') Corresponding immunoblot using mAb AND-1/23-5-14. R, reference proteins, see Fig. 1 and lane 9. Fraction numbers are indicated on the top of the lanes (top of the gradient to the left), except lane 9 which contains reference proteins (as in lane R). Bars on the top of the lanes indicate the peak positions of the reference proteins BSA (4.3 S; fractions 3 and 4), catalase (11.3 S; fractions 7 and 8) and thyroglobulin (16.5 S; fractions 11 and 12). The bulk of protein AND-1 is recovered in fractions 5 and 6, i.e. with a sedimentation coefficient of ~7.5 S, which is indicative of an oligomeric state of the protein. (B) Immunoblot of proteins from XLKE-A6 cell extracts fractionated by gel filtration and separated by SDS-PAGE. Bars on the top of the lanes indicate the peak positions of the co-fractionated reference proteins: thyroglobulin (M_{app} , 669,000; fractions 7 and 8), ferritin (M_{app} , 440,000; fraction 11) and catalase (M_{app} , 232,000; fraction 15). Protein AND-1 was detected mainly in fractions 5-7, corresponding to M_{app} 800,000, indicative of an oligomeric, probably dimeric state of protein AND-1.

corresponding region of the SepB protein is shown in Fig. 7A. The sequence homology extends beyond the mere occurrence of the repeats. Although the primary structure of WD-repeats is not well conserved, they have a conserved secondary structure. A WD-repeat was defined as a β -strand-turn- β -strand-turn- β -strand alternation (Neer et al., 1994). This was also predicted for the WD-repeats found in protein AND-1 (Fig. 7B).

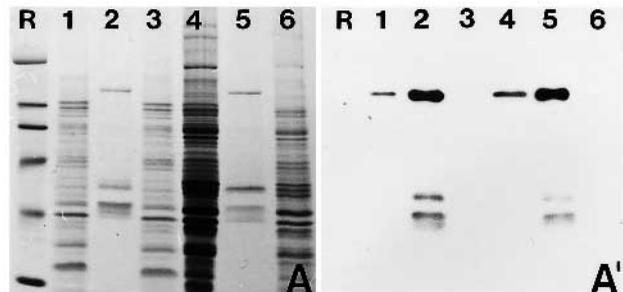


Fig. 4. Immunoprecipitation of protein AND-1. (A) Proteins of the different fractions separated by SDS-PAGE and stained with Coomassie blue. (A') Corresponding immunoblot using mAb AND-1/23-5-14. R, reference proteins, see Fig. 1. Lane 1, proteins of the *Xenopus* oocyte nuclei HSS fraction prior to immunoprecipitation. Lane 2, immunoprecipitate obtained from HSS-material. Lane 3, supernatant fraction corresponding to lane 2. Lane 4, proteins of *Xenopus* egg extract prior to immunoprecipitation. Lane 5, immunoprecipitate obtained from the egg extract. Lane 6, supernatant fraction corresponding to lane 5. Protein AND-1 was immunoprecipitated in Coomassie blue-visible amounts (A, lanes 2 and 5), but no co-precipitating protein was detectable. The lower bands in lane 2 and 5 represent the heavy IgG chains of mAb AND-1/23-5-14.

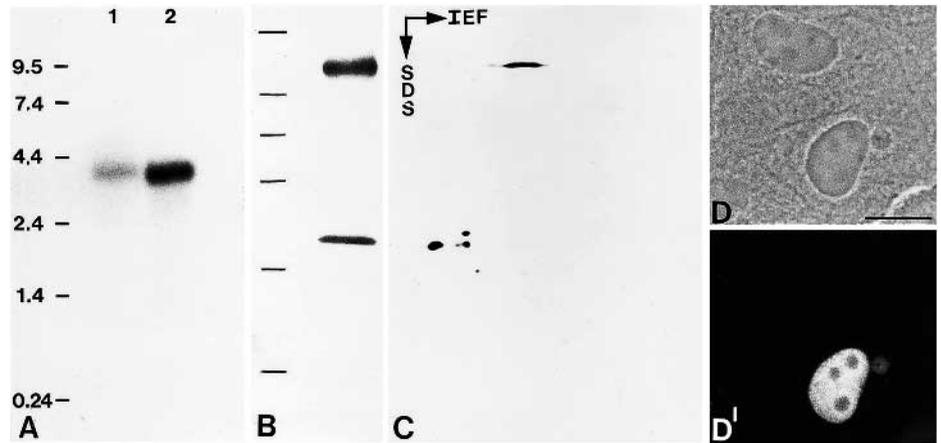
A second conspicuous feature of protein AND-1 is the presence of a single characteristic HMG-box in the carboxy-terminal part of the protein (aa positions 1013-1076), indicating that this protein belongs to the HMG-box superfamily of DNA-binding proteins (Fig. 7C) involved, for example, in regulation of chromatin formation, replication and transcription, (Johns, 1982; Jantzen et al., 1990; Griess et al., 1993; Laudet et al., 1993). An HMG-box is defined as three α -helices able to adopt an L-shaped arrangement (Baxevanis et al., 1995; Teo et al., 1995). The secondary structure prediction of the AND-1 HMG box showed the same helical composition (Fig. 7C).

Protein AND-1 also contains some clusters of acidic amino acids (boxed in Fig. 6A), as they have been described in various other major nuclear proteins (e.g. Kleinschmidt et al., 1986; Schmidt-Zachmann et al., 1987; Schmidt-Zachmann and Franke, 1988; for review see Earnshaw, 1987). The overall structural organization of protein AND-1 is schematically shown in Fig. 6B.

We also noticed candidates for nuclear localization signals (NLS; for review see Boulikas, 1993) within the amino acid sequence of the protein (underlined in Fig. 6A). Griess et al. (1993) and Poulat et al. (1995) have proposed a bipartite NLS within each HMG-box motif, based on the NLS sequence requirements discussed by Dingwall and Laskey (1991). In protein AND-1, this might correspond to KKPX₁₀R (aa 1,013-1,027). Other possible NLS sequences might be KKKKX₁₁RK (aa 599-615), somewhat reversely similar to the bipartite NLS of *Xenopus* nucleoplamin (KR_{X10}KKKK; Robbins et al., 1991) or KKRKR (aa 1,078-1,082), similar to the monopartite NLS of human lamin A (KKRK; Loewinger and McKeon, 1988).

The sequence similarities of protein AND-1 to both the WD-repeat and the HMG-box families are restricted to these motifs, with the single exception of protein SepB of *Aspergillus nidulans*. This protein of 710 amino acids contains WD-repeats in its amino-terminal part but is 417 amino acids shorter,

Fig. 5. Molecular characterization of the cDNA clone encoding protein AND-1. (A) Identification of the AND-1 mRNA by northern blot analysis. RNA from *Xenopus* oocytes was separated, transferred to a membrane and probed with antisense cRNA derived from pBS-AND-1-*Xen*. Lane 1, resulting autoradiograph after probing 20 µg total RNA. Lane 2, resulting autoradiograph using 5 µg poly(A)⁺ RNA. The positions of different RNAs (kb values are given on the left) used for reference are indicated by bars. Note the appearance of a single band corresponding to a ~4 kb mRNA. (B) Autoradiograph of the [³⁵S]methionine-labeled AND-1 translation products. The AND-1 cDNA was subcloned into the mammalian expression vector pRc/CMV and used in a coupled in vitro transcription/translation reaction. The resulting translation products were separated by SDS-PAGE. Horizontal bars on the left indicate the position of reference proteins (see legend to Fig. 1). (C) Autoradiograph showing the same translation products separated by two-dimensional gel electrophoresis (iso-electric focussing/SDS-PAGE). Note the identical SDS-PAGE mobility and apparent pI of the major in vitro translation product as compared to native protein AND-1. The smaller signals probably represent degradation products. (D) Phase-contrast optics of cultured human hepatocellular carcinoma cells of line PLC used to determine the subcellular localization of the protein encoded by pRc/CMV-AND-1-*Xen* in transfection experiments. (D') Corresponding epifluorescence using the *Xenopus*-specific mAb AND-1/23-5-14. Transfected PLC cells showed a strong nucleoplasmic staining, not distinguishable from native AND-1 protein in XLKE-A6 cells (for comparison see Fig. 2). Bar, 20 µm.



lacking the entire region homologous to the HMG-box-containing portion of protein AND-1 (cf. Fig. 7A), and shows 22.9% identity (50.8% similarity) with protein AND-1.

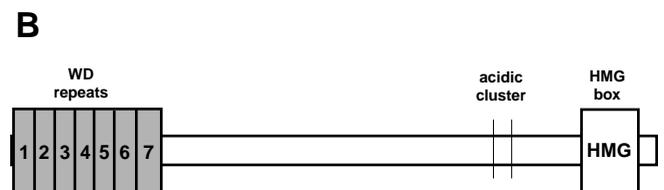
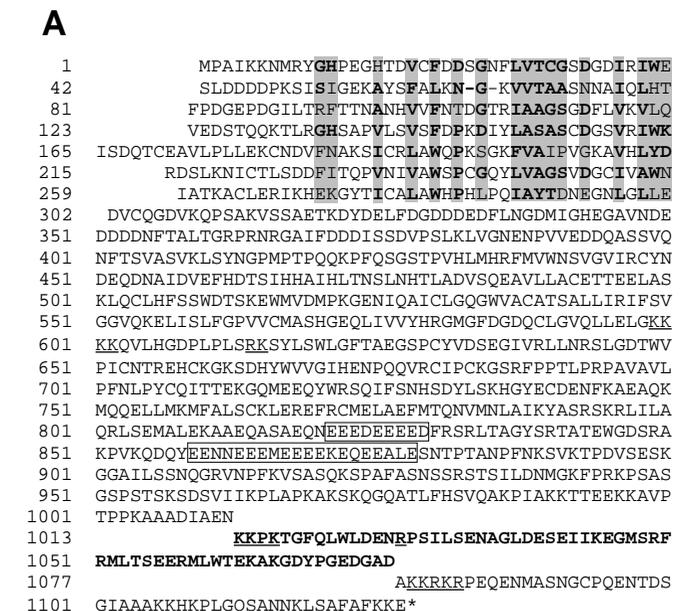
Binding of protein AND-1 to DNA

The DNA-binding capacity of the native protein AND-1 was tested by DNA-affinity chromatography, the classic approach used to enrich DNA-binding proteins (Alberts et al., 1968) and to purify transcription factors (Rosenfeld and Kelly, 1986). The soluble proteins of the oocyte nuclear HSS-fraction were applied to native, double-stranded DNA coupled to cellulose, the DNA-binding proteins were eluted at increasing salt concentrations, and the fractions obtained were analyzed by immunoblotting experiments (Fig. 8A,A'). The protein was found to be bound to DNA quantitatively and with high affinity. It was completely eluted from the column using buffer containing 150 mM NaCl (Fig. 8A' lane 4). Using the same assay, we could also prove the DNA-binding capacity of the recombinant expressed AND-1 HMG-box (Fig. 8B), indicating that

native protein AND-1 present in the HSS-fraction was bound directly to DNA via its HMG-box and not retained through a possible binding partner present in the HSS-fraction (results not shown).

It has been shown for several HMG-box proteins that they

Fig. 6. Amino acid sequence of protein AND-1. (A) The open reading frame of 1127 aa was deduced from the full-length AND-1 cDNA clone consisting of 3,978 nucleotides. The first 301 amino acids of protein AND-1 contain seven WD repeats. The consensus-sequence of WD repeats (for review see Neer et al., 1994) is restricted to the grey-shadowed positions; matching amino acids are highlighted in bold letters (see also Fig. 7A). The HMG box found in the carboxy-terminal part of AND-1 (aa 1013-1076) is marked with bold letters. Two stretches of acidic amino acids are denoted by boxes, and putative nuclear localization signals (NLS) of AND-1 are underlined. (B) Scheme of the structural organization of protein AND-1. The seven WD repeats at the amino terminus of AND-1 are represented by numbered grey boxes. The HMG-box in the carboxy-terminal part of AND-1 is denoted by a thick-lined box. The vertical dashes represent acidic stretches within the amino acid sequence of protein AND-1.



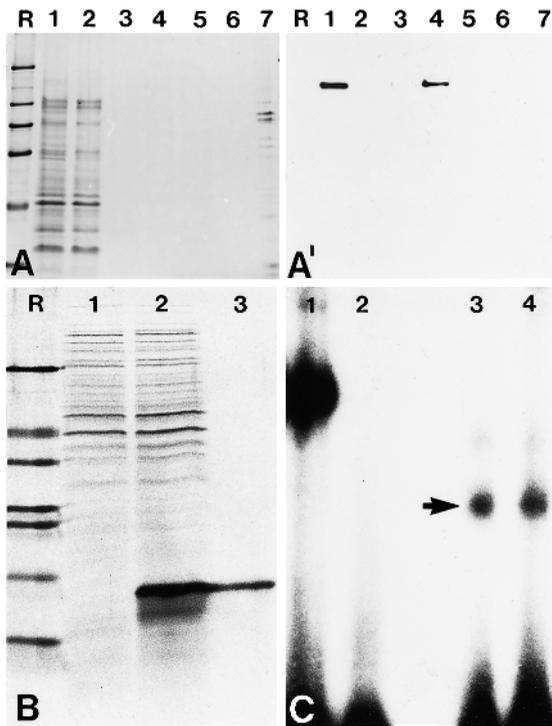


Fig. 8. AND-1 is a DNA-binding protein. (A) Soluble nuclear proteins of the HSS-fraction from *Xenopus* oocyte nuclei were applied to a column of native, double-stranded DNA coupled to cellulose. Bound proteins were eluted with buffers of ascending salt concentrations, collected, separated by SDS-PAGE and stained with Coomassie blue. (A') Corresponding immunoblot using mAb AND-1/23-5-14. R, reference proteins, see Fig. 1. Lane 1, proteins of the HSS fraction, prior to loading on the column. Lane 2, flow-through fraction. Lanes 3-7, bound proteins were eluted with buffer containing 100 mM NaCl (lane 3), 150 mM NaCl (lane 4), 200 mM (lane 5), 250 mM (lane 6) and 1.5 M NaCl (lane 7). The acidic protein AND-1 recognized by mAb AND-1/23-5-14 was bound completely to the column, indicating its high affinity to DNA. It could be eluted with buffer containing 150 mM NaCl. (B) Expression of the DNA-binding motif of protein AND-1. An AND-1 fragment (aa 1,007-1,106) containing the HMG-box (aa 1,013-1,076) was expressed in *E. coli* and subsequently purified. The proteins were separated by SDS-PAGE and stained with Coomassie blue. R, reference proteins (from top to bottom in kDa): 66, 45, 36, 29, 24, and 14.2. Lane 1, 20 μ l samples of total lysed *E. coli* containing the expression vector before induction. Lane 2, same amount of lysed cells containing the expression vector 3 hours after induction. Lane 3, the recombinant AND-1 HMG-box after purification on a Ni-NTA column. (C) Binding of the AND-1 HMG-box to four-way junction DNA. Lane 1, the shift of four-way junction DNA was controlled using 20 ng recombinant mouse UBF-1 expressed in the baculovirus system. Lane 2, labeled four-way junction DNA without added protein. Lanes 3,4, mobility shift assay using 600 ng (lane 3) and 400 ng (lane 4) of the bacterial-expressed HMG-box of protein AND-1. The four-way junction DNA shifted by the AND-1 HMG-box is marked by an arrow.

homologue in rat, we examined cDNAs derived from rat PC12 cells, brain and liver, in comparison with *Xenopus* ovary, kidney and liver, by using PCR with two oligonucleotides corresponding to the AND-1 cDNA sequence. In all cases, the correct PCR-segment of 400 nt could be amplified (data not shown). A similar result was obtained with cDNAs from

human brain, epidermis, liver and stomach by PCR, suggesting that mRNA encoding protein AND-1 also occurs in mammalian species in very different cells and tissues.

DISCUSSION

The large (125 kDa) nucleoplasmic protein AND-1 that we have discovered in oocytes and various other cells of *Xenopus laevis* is a remarkable case of a naturally occurring chimeric molecule combining hallmark features of two different families of regulatory proteins, the WD-repeat and the HMG-box proteins, suggesting that the DNA-binding protein AND-1 is also involved in protein-protein interactions. A highly related protein has also been identified in various mammals, and a partial cDNA encoding the orthologous human protein has recently been cloned from somatic cells (results not shown). Therefore, we conclude that the gene encoding this protein has been evolutionarily conserved and is widely, if not ubiquitously, expressed in different cell types and under different conditions.

Within the superfamily of HMG-box proteins, protein AND-1 falls into the subfamily of those members which contain only a single box. These proteins recognize, and bind to, DNA with high sequence specificity (Laudet et al., 1993) and rapidly and effectively accumulate in a soluble form dispersed throughout the nucleoplasm (for reviews see Landsmann and Bustin, 1993; Laudet et al., 1993), in contrast to other HMG-box proteins which, in the same cells, distribute over both compartments, the nucleoplasm and the cytoplasm, including the prototype proteins HMG1 and HMG2 (e.g. Bustin and Neihart, 1979; Kleinschmidt et al., 1983a; Einck and Bustin, 1985; Landsmann and Bustin, 1993 and further references therein). Several HMG-box proteins of both subfamilies, whether they bind to DNA in a sequence-specific or -unspecific manner, also recognize non-B-DNA structures like four-way junctions, which can be generated by recombination events and by intrastrand base pairing of inverted repeat sequences. Obviously, this is also true for protein AND-1, since the binding of the recombinant AND-1 HMG-box to such structures has been demonstrated in the electrophoretic mobility shift assay.

Our observations that diffusible homodimer complexes of protein AND-1 accumulate in the interchromatin nucleoplasm and that the protein appears to be absent from nucleoli and dense chromatin, including mitotic chromosomes, argue against a specific function in inactivated genes and in rDNA transcription, in contrast to the HMG-box-containing UBF proteins (see, e.g. Jantzen et al., 1990; Zatsepina et al., 1993) and the single-strand nucleic acid-binding *Drosophila* protein, DssRP (Hsu et al., 1993). On the other hand, the occurrence of protein AND-1 in many diverse cell types indicates that its function is not exclusive to a given cell type or developmental stage. In this respect, the expression pattern of protein AND-1 is fundamentally different from those of many other HMG-box proteins that occur only in certain cells, such as the sex-determining proteins of the SRY-Sox group (e.g. Gubbay et al., 1990; Sinclair et al., 1990; Laudet et al., 1993; Harley and Goodfellow, 1994), the lymphocyte differentiation-specific transcription factors LEF-1 and TCF-1 (e.g. Travis et al., 1991; Van de Wetering et al., 1991), and the stage-specific POP-1

protein of *C. elegans* (Lin et al., 1995). The occurrence of protein AND-1 in the fraction containing the 'soluble' nucleoplasmic proteins also shows that it is not non-specifically associated with chromatin, but suggests that its reaction from the soluble to the DNA-bound is regulated.

As first shown for the specific binding of the multiple HMG-box-containing protein UBF to one of the RNA polymerase I subunits (Schnapp et al., 1994), HMG-box motifs not only effect DNA recognition and binding but can also determine protein-protein interactions. A special form of such interaction appears to be the homodimer formation shown to involve an HMG-box in UBF (O'Mahony et al., 1992) or in protein HMG1 (Bianchi et al., 1992). We are currently testing whether the dimer formation of protein AND-1 observed in *Xenopus* oocytes and eggs is also mediated by the single HMG-box present.

Protein-protein interaction as a major reaction and function has also been ascribed to WD-repeats, the other family-defining motif prominent in protein AND-1. Proteins possessing variable numbers of WD-repeats of around 40 amino acids have been found in eukaryotes (Neer et al., 1994) and, recently, also in one prokaryote (Janda et al., 1996) and have been reported to determine the local protein conformation and to contribute to a diversity of nuclear and cytoplasmic functions (reviewed by Neer et al., 1994; Neer and Smith, 1996). Such functions include signal transductions (for review see Simon et al., 1991; Neer, 1995), histone binding, chaperoning and escorting (Kaufman et al., 1995; Edmondson et al., 1996), chromatin replication or control of cell cycle and mitosis (Yochem and Byers, 1987; Kaufman et al., 1995; Muro et al., 1995), chromosome segregation and cytokinesis (Harris and Hamer, 1995), gene transcription and suppression (e.g. Keleher et al., 1992; Dynlacht et al., 1993; Koleske and Young, 1994; Komachi et al., 1994; Paroush et al., 1994; Gutjahr et al., 1995; Sathe and Harte, 1995), mRNA modification (Takagaki and Manley, 1992), nuclear pore complex architecture (Siniosoglou et al. 1996), and also vesicle formation and fusion (Harrison-Lavoie et al., 1993; Pryer et al., 1993; Stenbeck et al., 1993; Gerich et al., 1995). Moreover, defects in some WD-repeat proteins have been discussed in relation to appearances of specific inheritable developmental disorders (e.g. Henning et al., 1995; Lorain et al., 1996).

Among the members of this family, protein AND-1 has a 'normal' number of repeats (7) of a normal length distribution, but stands out by its relatively large size (see, however, similarly-sized polypeptides such as α - and β '-COPs; Gerich et al., 1995; Stenbeck et al., 1993; Harrison-Lavoie et al., 1993) and its generally nucleoplasmic location (for some WD-repeat proteins with nuclear enrichment, dependent on the specific cell cycle or developmental stages; see Gutjahr et al., 1995; Muro et al., 1995). The special relationship of protein AND-1 to protein SepB of *Aspergillus nidulans*, which extends beyond the WD-repeat region, points to the possibility of a functionally close relationship to this fungal protein, which is known to be involved in the regulation of chromosome segregation and cytokinesis (Harris and Hamer, 1995).

At present, the functional significance of the WD-repeat domain with respect to these diverse functions is still not fully understood. The prevailing general concept that WD-repeats might be directly involved in, and essential for, specific protein-protein interactions, including examples in which they

act as assembly-promoting factors or chaperons (Edmondson et al., 1996; Kaufman et al., 1995; Sathe and Harte, 1995; Komachi et al., 1994; Neer et al., 1994; Conklin and Bourne, 1993; Keleher et al., 1992; van der Voorn and Ploegh, 1992) seems to be challenged by demonstrations that in certain proteins the entire domain containing these repeats is not needed for effecting the functionally important binding to another protein (e.g. Paroush et al., 1994). We are currently searching for possible AND-1-binding proteins to learn more about the functional mechanisms involving this complex 'dual family membership' nuclear protein.

We thank Sonja Reidenbach, Cilly Kuhn and Michaela Hergt for expert technical assistance, Jutta Müller-Osterholt for preparation of photographs and Peter Eichhorn for excellent care of the animals. We gratefully acknowledge Dr Renate Voit (this center, Division of Molecular Biology of the Cell II) for valuable discussions and for the help with the electrophoretic mobility shift assay, and Rudolf F. Zirwes for valuable discussions and for kindly providing XLKE-A6 cell-extracts fractionated by gel filtration. This work has been supported in parts by the Deutsche Forschungsgemeinschaft (grant Schm 862/2-3 to M.S.-Z.).

REFERENCES

- Alberts, B. M., Amodio, F. J., Jenkins, M., Gutmann, E. D. and Ferris, F. L. (1968). Studies with DNA-cellulose chromatography. I. DNA-binding proteins from *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 289-305.
- Ankenbauer, T., Kleinschmidt, J. A., Walsh, M. J., Weiner, O. H. and Franke, W. W. (1989). Identification of a widespread nuclear actin binding protein. *Nature* **342**, 822-825.
- Bachvarov, D. and Moss, T. (1991). The RNA polymerase I transcription factor xUBF contains 5 tandemly repeated HMG homology boxes. *Nucl. Acids Res.* **19**, 2331-2335.
- Baxevasis, A. D., Bryant, S. H. and Landsman, D. (1995). Homology model building of the HMG-1 box structural domain. *Nucl. Acids Res.* **23**, 1019-1029.
- Bianchi, M. E. (1988). Interaction of a protein from rat liver nuclei with cruciform DNA. *EMBO J.* **7**, 843-849.
- Bianchi, M. E., Falciola, L., Ferrari, S. and Lilley, D. M. (1992). The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. *EMBO J.* **11**, 1055-1063.
- Boulikas, T. (1993). Nuclear localization signals (NLS). *Crit. Rev. Euk. Gene Exp.* **3**, 193-227.
- Bruhn, S. L., Pil, P. M., Essigmann, J. M., Housman, D. E. and Lippard, S. J. (1992). Isolation and characterization of human cDNA clones encoding a high mobility group box protein that recognizes structural distortions to DNA caused by binding of the anticancer agent cisplatin. *Proc. Nat. Acad. Sci. USA* **89**, 2307-2311.
- Busch, S. J. and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet.* **6**, 36-40.
- Bustin, M. and Neihart, N. K. (1979). Antibodies against chromosomal HMG proteins stain the cytoplasm of mammalian cells. *Cell* **16**, 181-189.
- Chen, C. and Okayama, H. (1987). High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell Biol.* **7**, 2745-2752.
- Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.
- Conklin, B. R. and Bourne, H. R. (1993). Structural elements of G alpha subunits that interact with G beta gamma, receptors, and effectors. *Cell* **73**, 631-641.
- Cordes, V. C., Reidenbach, S., Köhler, A., Stuurman, N., van Driel, R. and Franke, W. W. (1993). Intranuclear filaments containing a nuclear pore complex protein. *J. Cell Biol.* **123**, 1333-1344.
- Dingwall, C. and Laskey, R. A. (1991). Nuclear targeting sequences - a consensus? *Trends Biochem. Sci.* **16**, 478-481.
- Dumont, J. N. (1972). Oogenesis in *Xenopus laevis* (Daudin). I. Stages of

- oocyte development in laboratory maintained animals. *J. Morphol.* **136**, 153-179.
- Dynlacht, B. D., Weinzierl, R. O., Admon, A. and Tjian, R.** (1993). The dTAFII80 subunit of Drosophila TFIID contains beta-transducin repeats. *Nature* **363**, 176-179.
- Earnshaw, W. C.** (1987). Anionic regions in nuclear proteins. *J. Cell Biol.* **105**, 1479-1482.
- Edmondson, D. G., Smith, M. M. and Roth, S. Y.** (1996). Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. *Genes Dev.* **10**, 1247-1259.
- Einck, L. and Bustin, M.** (1985). The intracellular distribution and function of the high mobility group chromosomal proteins. *Exp. Cell Res.* **156**, 295-310.
- Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell Badge, R. and Bianchi, M. E.** (1992). SRY, like HMGI, recognizes sharp angles in DNA. *EMBO J.* **11**, 4497-4506.
- Gerich, B., Orci, L., Tschochner, H., Lottspeich, F., Ravazzola, M., Amherdt, M., Wieland, F. and Harter, C.** (1995). Non-clathrin-coat protein alpha is a conserved subunit of coatamer and in *Saccharomyces cerevisiae* is essential for growth. *Proc. Nat. Acad. Sci. USA* **92**, 3229-3233.
- Griess, E. A., Rensing, S. A., Grasser, K. D., Maier, U. G. and Feix, G.** (1993). Phylogenetic relationships of HMG box DNA-binding domains. *J. Mol. Evol.* **37**, 204-210.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell Badge, R.** (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245-250.
- Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R. A. and Noll, M.** (1995). The Polycomb-group gene, extra sex combs, encodes a nuclear member of the WD-40 repeat family. *EMBO J.* **14**, 4296-4306.
- Harley, V. R. and Goodfellow, P. N.** (1994). The biochemical role of SRY in sex determination. *Mol. Reprod. Dev.* **39**, 184-193.
- Harris, S. D. and Hamer, J. E.** (1995). sepB: an *Aspergillus nidulans* gene involved in chromosome segregation and the initiation of cytokinesis. *EMBO J.* **14**, 5244-5257.
- Harrison, S. C.** (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715-719.
- Harrison-Lavoie, K. J., Lewis, V. A., Hynes, G. M., Collison, K. S., Nutland, E. and Willison, K. R.** (1993). A 102 kDa subunit of a Golgi-associated particle has homology to beta subunits of trimeric G proteins. *EMBO J.* **12**, 2847-2853.
- Heid, H. W., Schmidt, A., Zimbelmann, R., Schäfer, S., Winter-Simanowski, S., Stumpp, S., Keith, M., Figge, U., Schnölzer, M. and Franke, W. W.** (1994). Cell type-specific desmosomal plaque proteins of the plakoglobin family: plakophilin 1 (band 6 protein). *Differentiation* **58**, 113-131.
- Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., Schultz, R. A., Stefanini, M., Lehmann, A. R., Mayne, L. V. and et al.** (1995). The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIF. *Cell* **82**, 555-564.
- Hsu, T., King, D. L., LaBonne, C. and Kafatos, F. C.** (1993). A Drosophila single-strand DNA/RNA-binding factor contains a high-mobility-group box and is enriched in the nucleolus. *Proc. Nat. Acad. Sci. USA* **90**, 6488-6492.
- Hügler, B., Kleinschmidt, J. A. and Franke, W. W.** (1983). The 22 S cylinder particles of *Xenopus laevis*. II. Immunological characterization and localization of their proteins in tissues and cultured cells. *Eur. J. Cell Biol.* **32**, 157-163.
- Hügler, B., Scheer, U. and Franke, W. W.** (1985). Ribocharin: a nuclear Mr 40,000 protein specific to precursor particles of the large ribosomal subunit. *Cell* **41**, 615-627.
- Janda, L., Tichy, P., Spizek, J. and Petricek, M.** (1996). A deduced *Thermoplasma curvatum* protein containing serine/threonine protein kinase and WD-repeat domains. *J. Bacteriol.* **178**, 1487-1489.
- Jantzen, H. M., Admon, A., Bell, S. P. and Tjian, R.** (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344**, 830-836.
- Johns, E. W.** (1982). *The HMG Chromosomal Proteins*. Academic Press, Inc., New York.
- Kaufman, P. D., Kobayashi, R., Kessler, N. and Stillman, B.** (1995). The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105-1114.
- Keleher, C. A., Redd, M. J., Schultz, J., Carlson, M. and Johnson, A. D.** (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. *Cell* **68**, 709-719.
- Kleinschmidt, J. A. and Franke, W. W.** (1982). Soluble acidic complexes containing histones H3 and H4 in nuclei of *Xenopus laevis* oocytes. *Cell* **29**, 799-809.
- Kleinschmidt, J. A., Scheer, U., Dabauvalle, M. C., Bustin, M. and Franke, W. W.** (1983a). High mobility group proteins of amphibian oocytes: a large storage pool of a soluble high mobility group-1-like protein and involvement in transcriptional events. *J. Cell Biol.* **97**, 838-848.
- Kleinschmidt, J. A., Hügler, B., Grund, C. and Franke, W. W.** (1983b). The 22 S cylinder particles of *Xenopus laevis*. I. Biochemical and electron microscopic characterization. *Eur. J. Cell Biol.* **32**, 143-156.
- Kleinschmidt, J. A., Fortkamp, E., Krohne, G., Zentgraf, H. and Franke, W. W.** (1985). Co-existence of two different types of soluble histone complexes in nuclei of *Xenopus laevis* oocytes. *J. Biol. Chem.* **260**, 1166-1176.
- Kleinschmidt, J. A., Dingwall, C., Maier, G. and Franke, W. W.** (1986). Molecular characterization of a karyophilic, histone-binding protein: cDNA cloning, amino acid sequence and expression of nuclear protein N1/N2 of *Xenopus laevis*. *EMBO J.* **5**, 3547-3552.
- Koleske, A. J. and Young, R. A.** (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466-469.
- Komachi, K., Redd, M. J. and Johnson, A. D.** (1994). The WD repeats of Tup1 interact with the homeo domain protein alpha 2. *Genes Dev.* **8**, 2857-2867.
- Köhler, G. and Milstein, C.** (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497.
- Krohne, G. and Franke, W. W.** (1980a). Immunological identification and localization of the predominant nuclear protein of the amphibian oocyte nucleus. *Proc. Nat. Acad. Sci. USA* **77**, 1034-1038.
- Krohne, G. and Franke, W. W.** (1980b). A major soluble acidic protein located in nuclei of diverse vertebrate species. *Exp. Cell Res.* **129**, 167-189.
- Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M. and Grummt, I.** (1994). Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity. *EMBO J.* **13**, 416-424.
- Landsman, D. and Bustin, M.** (1993). A signature for the HMG-1 box DNA-binding proteins. *BioEssays* **15**, 539-546.
- Laudet, V., Stehelin, D. and Clevers, H.** (1993). Ancestry and diversity of the HMG box superfamily. *Nucl. Acids Res.* **21**, 2493-2501.
- Lee, N. H., Weinstock, K. G., Kirkness, E. F., Earle Hughes, J. A., Fuldner, R. A., Marmaros, S., Glodek, A., Gocayne, J. D., Adams, M. D., Kerlavage, A. R. et al.** (1995). Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *Proc. Nat. Acad. Sci. USA* **92**, 8303-8307.
- Leube, R. E., Leimer, U., Grund, C., Franke, W. W., Harth, N. and Wiedenmann, B.** (1994). Sorting of synaptophysin into special vesicles in nonneuroendocrine epithelial cells. *J. Cell Biol.* **127**, 1589-1601.
- Lin, R., Thompson, S. and Priess, J. R.** (1995). pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early *C. elegans* embryos. *Cell* **83**, 599-609.
- Loewinger, L. and McKeon, F.** (1988). Mutations in the nuclear lamin proteins resulting in their aberrant assembly in the cytoplasm. *EMBO J.* **7**, 2301-2309.
- Lorain, S., Demczuk, S., Lamour, V., Toth, S., Aurias, A., Roe, B. A. and Lipinski, M.** (1996). Structural organization of the WD repeat protein-encoding gene HIRA in the DiGeorge Syndrome critical region of human chromosome 22. *Genome Res.* **6**, 43-50.
- Martinez-Balbás, M. A., Dey, A., Rabindran, S. K., Ozato, K. and Wu, C.** (1995). Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83**, 29-38.
- Miller, M., Kloc, M., Reddy, B., Eastman, E., Dreyer, C. and Etkin, L.** (1989). xlgv7: a maternal gene product localized in nuclei of the central nervous system in *Xenopus laevis*. *Genes Dev.* **3**, 572-583.
- Muro, Y., Chan, E. K., Landberg, G. and Tan, E. M.** (1995). A cell-cycle nuclear autoantigen containing WD-40 motifs expressed mainly in S and G₂ phase cells. *Biochem. Biophys. Res. Commun.* **207**, 1029-1037.
- Neer, E. J.** (1995). Heterotrimeric G proteins: organizers of transmembrane signals. *Cell* **80**, 249-257.
- Neer, E. J., Schmidt, C. J., Nambudripad, R. and Smith, T. F.** (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 297-300.

- Neer, E. J. and Smith, T. F. (1996). G protein heterodimers: new structures propel new questions. *Cell* **84**, 175-178.
- Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M. et al. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**, 75-80.
- O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021.
- O'Mahony, D. J., Smith, S. D., Xie, W. and Rothblum, L. I. (1992). Analysis of the phosphorylation, DNA-binding and dimerization properties of the RNA polymerase I transcription factors UBF1 and UBF2. *Nucl. Acids Res.* **20**, 1301-1308.
- Paroush, Z., Finley, R. L., Jr., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R. and Ish-Horowitz, D. (1994). Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**, 805-815.
- Peters, J. M., Walsh, M. J. and Franke, W. W. (1990). An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. *EMBO J.* **9**, 1757-1767.
- Peters, J. M., Franke, W. W. and Kleinschmidt, J. A. (1994). Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *J. Biol. Chem.* **269**, 7709-7718.
- Poulat, F., Girard, F., Chevron, M. P., Goze, C., Rebillard, X., Calas, B., Lamb, N. and Berta, P. (1995). Nuclear localization of the testis determining gene product SRY. *J. Cell Biol.* **128**, 737-748.
- Pryer, N. K., Salama, N. R., Schekman, R. and Kaiser, C. A. (1993). Cytosolic Sec13p complex is required for vesicle formation from the endoplasmic reticulum in vitro. *J. Cell Biol.* **120**, 865-875.
- Robbins, J., Dilworth, S. M., Laskey, R. A. and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasm nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615-623.
- Rosenfeld, P. J. and Kelly, T. J. (1986). Purification of nuclear factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* **261**, 1398-1408.
- Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. *Meth. Enzymol.* **266**, 525-539.
- Rost, B. and Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Nat. Acad. Sci. USA* **90**, 7558-7562.
- Sathe, S. S. and Harte, P. J. (1995). The Drosophila extra sex combs protein contains WD motifs essential for its function as a repressor of homeotic genes. *Mech. Dev.* **52**, 77-87.
- Saurin, A. J., Borden, K. L. B., Boddy, M. N. and Freemont, P. S. (1996). Does this have a familiar RING? *Trends Biochem. Sci.* **21**, 208-214.
- Scalenghe, F., Buscaglia, M., Steinheil, C. and Crippa, M. (1978). Large scale isolation of nuclei and nucleoli from vitellogenic oocytes of *Xenopus laevis*. *Chromosoma* **66**, 299-308.
- Schmidt-Zachmann, M. S. and Franke, W. W. (1988). DNA cloning and amino acid sequence determination of a major constituent protein of mammalian nucleoli. Correspondence of the nucleoplasm-related protein NO38 to mammalian protein B23. *Chromosoma* **96**, 417-426.
- Schmidt-Zachmann, M. S., Hügle, B., Scheer, U. and Franke, W. W. (1984). Identification and localization of a novel nucleolar protein of high molecular weight by a monoclonal antibody. *Exp. Cell Res.* **153**, 327-346.
- Schmidt-Zachmann, M. S., Hügle, B. and Franke, W. W. (1987). A constitutive nucleolar protein identified as a member of the nucleoplasm family. *EMBO J.* **6**, 1881-1890.
- Schnapp, G., Santori, F., Carles, C., Riva, M. and Grummt, I. (1994). The HMG box-containing nucleolar transcription factor UBF interacts with a specific subunit of RNA polymerase I. *EMBO J.* **13**, 190-199.
- Siegel, L. M. and Monty, K. J. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**, 346-362.
- Simon, M. I., Strathmann, M. P. and Gautam, N. (1991). Diversity of G proteins in signal transduction. *Science* **252**, 802-808.
- Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R. and Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-244.
- Siniosoglou, S., Wimmer, C., Rieger, M., Doye, V., Tekotte, H., Weise, C., Emig, S., Segref, A. and Hurt, E. C. (1996). A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell* **84**, 265-275.
- Staben, C. and Yanofsky, C. (1990). *Neurospora crassa* a mating-type region. *Proc. Nat. Acad. Sci. USA* **87**, 4917-4921.
- Stenbeck, G., Harter, C., Brecht, A., Herrmann, D., Lottspeich, F., Orci, L. and Wieland, F. T. (1993). beta'-COP, a novel subunit of coatamer. *EMBO J.* **12**, 2841-2845.
- Takagaki, Y. and Manley, J. L. (1992). A human polyadenylation factor is a G protein beta-subunit homologue. *J. Biol. Chem.* **267**, 23471-23474.
- Teo, S.-H., Grasser, K. D., Hardman, C. H., Broadhurst, R. W., Laue, E. D., and Thomas, J. O. (1995). Two mutations in the HMG-box with very different structural consequences provide insights into the nature of binding to four-way junction DNA. *EMBO J.* **14**, 3844-3853.
- Thomas, J. O. and Kornberg, R. D. (1975). An octamer of histones in chromatin and free in solution. *Proc. Nat. Acad. Sci. USA* **72**, 2626-2630.
- Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. *Genes Dev.* **5**, 880-894.
- van der Voorn, L. and Ploegh, H. L. (1992). The WD-40 repeat. *FEBS Lett.* **307**, 131-134.
- van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.* **10**, 123-132.
- Wagner, C. R., Hamana, K. and Elgin, S. (1992). A high-mobility-group protein and its cDNA from *Drosophila melanogaster*. *Mol. Cell Biol.* **12**, 1915-1923.
- Yochem, J. and Byers, B. (1987). Structural comparison of the yeast cell division cycle gene CDC4 and a related pseudogene. *J. Mol. Biol.* **195**, 233-245.
- Zatsepin, O. V., Voit, R., Grummt, I., Spring, H., Semenov, M. V. and Trendelenburg, M. F. (1993). The RNA polymerase I-specific transcription initiation factor UBF is associated with transcriptionally active and inactive ribosomal genes. *Chromosoma* **102**, 599-611.

(Received 6 August 1996 – Accepted 26 February 1997)