Chromosoma (2004) 113: 244–257
DOI 10.1007/s00412-004-0314-4

RESEARCH ARTICLE

Xin Sun · Jian Zhao · Karin Kylberg · Teresa Soop · Kevin Palka · Erik Sonnhammer · Neus Visa · Alla T. Alzhanova-Ericsson · Bertil Daneholt

Conspicuous accumulation of transcription elongation repressor hrp130/CA150 on the intron-rich Balbiani ring 3 gene

Received: 16 July 2004 / Revised: 19 August 2004 / Accepted: 19 August 2004 / Published online: 6 October 2004

Abstract Chromosomal puffs on the polytene chromosomes in the dipteran *Chironomus tentans* offer the possibility of comparing the appearance of RNA-binding proteins at different transcription sites. We raised a monoclonal antibody that recognized a 130 kDa protein, designated hrp130. Immunocytological analysis of isolated chromosomes showed that hrp130 is heavily accumulated in a specific puff, called Balbiani ring 3; only occasionally is hrp130 abundant in one or two additional puffs on other chromosomes. The immunolabeling was sensitive to RNase treatment, suggesting that hrp130 is associated with nascent ribonucleoproteins. As shown by immunoelectron microscopy hrp130 is distributed along the active BR3 genes. The full sequence of hrp130 was determined by cDNA cloning. The protein comprises 1028 amino acids and contains three WW domains in the N-terminal half and six FF domains in the C-terminal half of the molecule. The protein is conserved from *Caenorhabditis elegans* to mammals; the human homolog is known as the transcription elongation repressor CA150. We propose that the abundance of hrp130/CA150 in BR3 is connected with the exceptionally high level of splicing in this locus and that hrp130/CA150 adjusts the transcription rate to the numerous splicing events taking place along the gene to ensure proper splicing.

Communicated by U. Scheer

X. Sun · J. Zhao · K. Kylberg · T. Soop · K. Palka · N. Visa · A. T. Alzhanova-Ericsson · B. Daneholt (✉)
Department of Cell and Molecular Biology, Medical Nobel Institute, Karolinska Institutet,
17177 Stockholm, Sweden
e-mail: bertil.daneholt@cmb.ki.se

E. Sonnhammer
Center for Genomics and Bioinformatics, Karolinska Institutet,
17177 Stockholm, Sweden

Present address:
N. Visa
Department of Molecular Biology and Functional Genomics,
Stockholm University,
10691 Stockholm, Sweden

Introduction

In a given cell nucleus, thousands of genes are being transcribed by RNA polymerase II. Concomitant with transcription, proteins are added to the primary transcript to form a thin ribonucleoprotein (RNP) fibril, which constitutes the substrate on which various proteins or protein assemblies act to process the primary transcript further (Dreyfuss et al. 1993; Fakan 1994; Krecic and Swanson 1999; Daneholt 2001). Capping and often also splicing of the primary transcript occur before transcription is completed (Bentley 1999; Hirosse and Manley 2000). Moreover, the addition of a poly(A) tail is closely connected with the termination of the transcription process (Baurén et al. 1998; Proudfoot et al. 2002). The growing RNP fibril is even prepared for transport co-transcriptionally (Lei et al. 2001; Strässer et al. 2002). Thus, during the course of transcription the nascent RNP fibrils associate with the macromolecular assemblies responsible for capping, splicing, polyadenylation and transport. It has recently been shown in a large number of studies that the various processing and transport events are coupled to each other in a delicate network of interactions, further emphasizing the importance of the co-transcriptional RNP assembly in gene expression (Maniatis and Reed 2002).

The proteins forming the RNP fibril, mainly hnRNP proteins and SR proteins, are abundant and bound to most or all growing mRNA transcripts (Sommerville et al. 1978; Lacroix et al. 1985; Roth and Gall 1987; Matunis et al. 1993; Wurtz et al. 1996). Furthermore, many of the additional proteins taking part in post-transcriptional processing and transport are likely to be present on essentially all types of transcripts. Thus, by far the most RNA-binding proteins associated with the primary transcripts seem to appear on most or all transcripts. However, in studies of lampbrush chromosomes (Pinol-

Roma et al. 1989; Lacroix et al. 1985) and polytene chromosomes (Kabisch and Bautz 1983; Matunis et al. 1993; Wurtz et al. 1996) it has been shown that there are quantitative, and in some cases also qualitative, differences between the chromosomal distributions of individual RNA-binding proteins. There are even reports suggesting that there could be more-or-less transcript-specific RNA-binding proteins (e.g., Sommerville et al. 1978; Roth and Gall 1987), but none of the proteins has yet been properly identified. Presumably, such transcript-specific RNA-binding proteins could play a crucial role in the regulation of one or more of the sophisticated processes taking place co-transcriptionally, and selectively affect the expression of a single gene or a small subset of genes.

For several years, we have studied the distribution of a number of RNA-binding proteins along polytene chromosomes in the salivary glands of the dipteran *Chironomus tentans* (Daneholt 2001). These unusually big chromosomes show a large number of expanded regions, called chromosomal puffs, which represent the transcription sites. Due to the possibility of identifying and comparing many transcription sites along specific chromosomes, we have now used this experimental material to look for RNA-associated proteins with a more or less transcript-specific chromosomal distribution.

There are hundreds of chromosomal puffs along the four polytene chromosomes in the salivary glands of *C. tentans* (Pelling 1964). On chromosome IV, there are three puffs of exceptional size, designated Balbiani rings (BR1, BR2, and BR3), which have been extensively studied (Daneholt 2001). The two largest BRs, BR1 and BR2, contain genes that are 35–40 kb in size and encode large-sized secretory proteins of 1 MDa (Wieslander 1994). The genes contain only four minor introns, and the transcript is therefore only marginally reduced in size during maturation. The assembly and transport of the large BR mRNP particles can be followed by electron microscopy, and the fate of specific hnRNP proteins associated with the BR RNA has been revealed (e.g., Visa et al. 1996). The BR3 gene is 10.9 kb in size and contains 38 introns (Paulsson et al. 1990). The introns are spliced concomitant with transcription (Wetterberg et al. 1996), and the mature BR3 mRNA is 5.5 kb long and encodes a 185 kDa protein (Dignam and Case 1990; Paulsson et al. 1990). The transcription, assembly and splicing processes on the BR3 gene can also be visualized in the electron microscope (Wetterberg et al. 2001). However, the BR3 RNP particles are considerably smaller than the BR1 and BR2 particles and cannot be identified during transport through the nucleoplasm and into the cytoplasm.

In the present study, we have identified an RNA-associated protein, hrp130, which seems to be confined to a single chromosomal locus, BR3; only occasionally is it found in one or two additional, smaller puffs. Using cDNA cloning we have determined the amino acid sequence of hrp130 and found that hrp130 is evolutionarily conserved and an ortholog of the mammalian transcriptional repressor CA150. We demonstrate by immunoelectron microscopy that hrp130 appears along the transcriptionally active BR3 genes, and show that hrp130 can be released from the genes by RNase treatment. Thus, hrp130/CA150 is probably associated with the growing transcripts along the BR3 genes. This result is in agreement with the current hypothesis that CA150 is an elongation factor. We propose that the abundance of hrp130/CA150 in BR3 is functionally related to the exceptionally high level of splicing in BR3. As it is known that the transcription repressor CA150 can bind to the phosphorylated C-terminal domain of the polymerase as well as to the essential splicing factor SP1, we further suggest that hrp130/CA150 adjusts the elongation rate to the many splicing events taking place co-transcriptionally along the BR3 gene.

Materials and methods

Tissue culture conditions

Culture of *C. tentans* was as described (Lezzi et al. 1981). Salivary glands were isolated from fourth instar larvae. Tissue culture cells of *C. tentans* were cultivated at 24°C as previously described (Wyss 1982).

Preparation of *C. tentans* salivary gland extract

Larvae of *C. tentans* were treated with 0.1 mg/ml pilocarpine for 5 h to reduce the amount of secretion. Salivary glands were isolated and put in ice cold ZO medium. The glands were sedimented at 20,000 *g* for 6 min at 4°C. The medium was discarded and sample buffer (4% SDS, 0.02% bromophenol blue, 20% glycerol, 200 mM dithiothreitol, 100 mM TRIS-HCl, pH 6.8) with 1 mM phenylmethylsulfonyl fluoride (PMSF) was added. The glands were frozen in liquid nitrogen and homogenized with a pestle while vortexing. This procedure was repeated once.

Preparation of nuclear protein extracts from *C. tentans* tissue culture cells

Tissue culture cells of *C. tentans* were washed in PBS, resuspended in TNM (10 mM triethanolamine-HCl buffer, pH 7.0, 100 mM NaCl, 1 mM $MgCl_2$) containing 0.2% Nonidet P40 and 0.1 mM PMSF, homogenized in a glass tissue grinder with a tight-fitting pestle, and centrifuged at 2000 *g* for 5 min at 4°C. The pellet containing the nuclei was washed with TNM, sonicated, and centrifuged at 7000 *g* for 10 min at 4°C. The supernatant was called RNP extract I (Wurtz et al. 1996). The pellet was resuspended in TNM/NP40 and treated with RNase A (final concentration 100 μg/ml) for 10 min on ice. The sample was centrifuged at 7000 *g* for 10 min at 4°C. The supernatant was called RNP extract II.

Preparation of ssDNA-binding proteins

The RNP extracts I and II were mixed with 0.5 ml (wet volume) of single-stranded DNA (ssDNA)-agarose (Pinol-Roma et al. 1988). The affinity column was equilibrated with 100 mM NaCl in 50 mM sodium phosphate buffer, pH 7.4. Subsequently, the column was first washed with the same buffer and then with addition of 1 mg heparin (Sigma) per milliliter. Heparin-resistant proteins were eluted with 2 M NaCl in 50 mM sodium phosphate buffer, pH 7.4. Aliquots of the high-salt fraction were mixed with trichloroacetic acid (final concentration 35%). After being kept on ice for 30 min, the precipitated proteins were collected by centrifugation for 20 min at 7000 *g*. The pellet was washed with acetone containing 0.1% concentrated HCl, and then with acetone and dried.

Immunization of Balb/c mice

Immunization was carried out following standard procedures as described by Harlow and Lane (1988). Approximately 15–20 μg of immunoprecipitated proteins were mixed with complete Freund adjuvant and injected intraperitoneally into Balb/c mice. Three booster injections with the same amount of antigen in incomplete adjuvant were given at 2 week intervals. Ten days after the last boost, the sera of the immunized mice were tested by immunoblotting to detect antibodies against *C. tentans* proteins. The mouse that gave the strongest response received an intravenous injection of antigen in PBS without adjuvant.

Fusion and screening for suitable monoclonal antibodies

Three days after the last injection, the spleen of the immunized mouse was macerated, and the spleen cells were fused with mouse myeloma Sp20 cells in the presence of polyethylene glycol 4000 (Gibco). The fused cells were plated onto 96 well plates and cultured in Opti-MEM medium. The standard hypoxanthine–aminopterin–thymidine (HAT) selection procedure was applied to select the hybridoma clones (Harlow and Lane 1988). The medium from wells with growing cells was tested by immunoblotting against *C. tentans* nuclear extracts prepared as described by Wurtz et al. (1996). In order to select antibodies suitable for cytological experiments, the media were also tested by immunostaining of polytene chromosome squashes as described by Visa et al. (1996). Cells from the positive wells were cloned at least three times by limiting dilution. Monoclonal antibody (mAb) 2G8, which recognized a 130 kD protein and bound specifically to BR3, was chosen for further analysis.

Raising of polyclonal antibodies to p130

In order to raise polyclonal antibodies specific to the *C. tentans* p130 protein, we synthesized a 100 amino acid sequence from the p130 COOH-terminal. This sequence was fused to a histidine-tag and the protein product will be referred to as p130-His. The cDNA sequence was amplified by the polymerase chain reaction (PCR) using specific primers (Pharmacia Biotech). The previously purified cDNA phage was used as template in the PCR (Boehringer Mannheim). The PCR product was purified (Pharmacia PCR Product Prep kit) prior to restriction digest with *Nco*I and *Hin*dIII (MBI Fermenta) in buffer B (Promega). The expression plasmid pET 21d (Novagen) was purified using a Pharmacia Flexi Prep kit prior to digestion using the same enzymes as above. The digested plasmid and the PCR products were purified from a 2% low melting agarose gel (Seaplaque GTG) using Pharmacia PCR Product Prep kit. Then 0.1 pmol of the p50 fragment was used for ligations with 0.003 pmol pET21d fragment, using Ready to Go T4 Ligase (Pharmacia Biotech). The ligation mixture was incubated at 16°C overnight and further transformed into competent Nova Blue cells (Novagen). After heat shock at 42°C for 40 s the cells were grown at 37°C in SOC medium for 1 h, and further grown overnight at 37°C on LB-Broth agar plates with ampicillin. The positive clones were sequenced to confirm that the sequence was correct (see "DNA sequencing and sequencing analysis") and further transformed as above into the expression vector BL21(DE3) pLysS (Novagen). The clones were selected on LB-Broth agar plates with ampicillin and chloramphenicol.

The BL21(DE3)pLysS cells containing the plasmid were grown in LB-Broth with ampicillin and chloramphenicol. Expression of the fusion protein p130-His was induced by the addition of 1 mM isopropyl-1-thio-galactopyranoside (Boehringer Mannheim), and the bacteria were grown for 2 h before harvest, frozen in liquid nitrogen and stored at −70°C. The p130-His was purified in denaturing conditions. The bacterial extract was first incubated in 8 M urea, 0.1 M NaH_2PO_4, 0.01 M TRIS-HCl, pH 8.0 overnight on a rotating wheel at room temperature. The extract was sonicated and pelleted at 20,000 *g* for 30 min, and p130-His was purified from the supernatant using the Ni-NTA system (Qiagen) on a column. The supernatant was incubated with the Ni-NTA resin for 2.5 h prior to washing with the same buffer as above, pH 5.9, and eluted at pH 4.5. Analysis by imunoblotting and Coomassie staining showed that the eluted fraction contained pure p130-His. This fraction was concentrated using Centriplus MWCO 3 (Amicon) and the buffer was exchanged for PBS.

Rabbits were immunized intramuscularly in the back of the thigh muscle. The first immunization was carried out using 0.2 mg of purified antigen p130-His in 0.25 ml PBS, together with 0.25 ml Freund complete adjuvant. The subsequent boosters were performed once every 4 weeks, using 0.1 mg antigen in 0.25 ml PBS, together with Freund incomplete adjuvant. The rabbits were bled before

the first immunization (pre-immune serum) and after the third and fourth immunization. Fourteen weeks after the first immunization, the rabbits were killed and all the serum was collected (death serum).

Affinity purification of the polyclonal antibody

The His-tagged p130 peptide was coupled to NHS-activated Sepharose (4 Fast Flow column; Pharmacia). Six milligrams of peptide in 0.5 M NaCl was added to 2 ml resin and the mixture was rotated for 2 h at room temperature. The resin was collected following incubation with 1 M TRIS-HCl, pH 8.0, to block unoccupied binding sites. After washing the resin, 1 ml was mixed with 1 ml PBS 10× and 9.5 ml antiserum (final bleeding). Specific antibodies were allowed to bind to the His-tagged p130 peptide while rotating at 4°C for 2 h. After subsequent washes with PBS, PBS + 500 mM NaCl and 10 mM NaH_2PO_4 the antibodies were eluted with 100 mM glycine, pH 3.0. Collected fractions were immediately neutralized with 1 M Na_2HPO_4, pH 8.0. The fractions of interest were pooled and used for immunoblot analysis and immunostaining of isolated chromosomes.

SDS-polyacrylamide gel electrophoresis and immunoblot analysis

Proteins in the RNP extracts were precipitated with cold acetone, resuspended in sample buffer and separated by electrophoresis in a 7.5% polyacrylamide separating gel along with the salivary gland extract. After electrophoresis, the proteins were blotted onto transfer membranes (Immobilion PVDF, Millipore) by using a semi-dry transfer apparatus (Bio-Rad) according to the manufacturer's instructions. Membranes were blocked for 30 min in PBS containing 10% dry milk powder and incubated with 2G8 hybridoma supernatant or affinity purified polyclonal antibody, diluted 1:5000, for 1 h. The membranes were washed and incubated with alkaline phosphatase-conjugated goat anti-mouse IgM and goat anti-rabbit IgG, respectively, for another hour. After washing, the protein bands were visualized by adding the substrates nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl phosphate in AP buffer (5 mM $MgCl_2$, 0.1 M NaCl, 0.1 M TRIS-HCl, pH 9.5). The enzymatic reaction was stopped in 1 mM EDTA, 10 mM TRIS-HCl, pH 8.0.

Isolation of polytene chromosomes

The salivary glands of *C. tentans* were pre-fixed with 2% paraformaldehyde in TKM buffer (100 mM KCl, 1 mM $MgCl_2$ and 10 mM triethanolamine-HCl, pH 7.0) at room temperature for 1 min and then washed with cold TKM for 5 min. The salivary glands were treated for 1 min in TKM buffer containing 2% Nonidet (NP-40) and transferred into 0.025% NP-40 in TKM buffer. To release polytene chromosomes, the salivary glands were repeatedly sucked through a micropipette with a narrow opening (250 μm in diameter). The isolated polytene chromosomes were then transferred onto a siliconized slide. In order to get the chromosomes to stick to the surface of the slides, the 0.025% NP-40 solution was replaced by fresh TKM. All manipulations during the isolation of the chromosomes were performed on a cold microscope stage at close to 0°C. The isolated polytene chromosomes were post-fixed with 4% paraformaldehyde in TKM buffer for 30 min at room temperature and washed three times with fresh TKM buffer.

Immunostaining of isolated polytene chromosomes

For immunocytological analysis, the isolated chromosomes were blocked with 50 μl of 2% bovine serum albumin (BSA) in TKM for 30 min at room temperature in a humid chamber and incubated for 45 min with 40 μl of 2G8 hybridoma supernatant. The slides were washed 3×5 min in 0.1% Tween 20 in TKM and then incubated for 90 min with 50 μl of gold-conjugated secondary antibody (Jackson ImmunoResearch Laboratories; IgG, 6 nm in diameter) diluted 1:50 in TKM buffer containing 0.5% BSA. The specimens were washed three times for 5 min in TKM and three times with distilled water. An immunogold silver enhancement solution (IntenSEM, Amersham) was added onto the specimens. The specimens were incubated for 10–15 min at room temperature, rinsed in distilled water, mounted in 30% glycerol, and finally examined and photographed in a Zeiss light microscope. The same procedure was used for the polyclonal, affinity-purified anti-hrp130 antibody (diluted 1:5000), while a gold-conjugated goat anti-rabbit antibody was used as secondary antibody.

RNase A treatment of isolated chromosomes

Isolated polytene chromosomes were pre-fixed as described above, and upon isolation they were immediately incubated with 100 μg/ml RNase A in TKM buffer for 60 min at room temperature. The RNase A-treated chromosomes were then rinsed in TKM, post-fixed and incubated with antibodies. As primary antibody we used either our own monoclonal 2G8 antibody or an affinity-purified goat anti-Drosophila RNA polymerase IIo antibody (Morris et al. 1997). Appropriate gold-conjugated anti-IgG antibodies (anti-mouse and anti-goat antibodies, respectively) were applied as secondary antibodies.

Immunoelectron microscopy

Chromosome IVs were isolated and the immunoreaction was carried out as described under "Immunostaining of

isolated polytene chromosomes." The diameter of the gold particles conjugated to secondary antibody was 12 nm.

After labeling with the secondary antibody, the specimens were post-fixed in 2% glutaraldehyde in TKM. They were rinsed in TKM for 3×5 min, dehydrated in 90% ethanol and embedded in an Agar 100 resin. Chromosome IVs were sectioned and stained with saturated uranyl acetate followed by lead citrate as previously described (Björkroth et al. 1988). The specimens were examined and photographed in a Philips CM 120 microscope at 60 kV.

Isolation of cDNA clones

The hybridoma culture supernatant with the monoclonal antibody 2G8 was diluted 1:100 to screen a randomly primed λgt11 cDNA library using ProtoBlot Immunoscreening System (Promega); the cDNA library was made from salivary glands of *C. tentans*. Antibody-specific clones were purified and sequenced. One of the clones was used as template for PCR amplification with λgt11 forward and reverse primers (Promega Biotech). The amplified DNA fragment was purified by using a Wizard PCR Preps DNA Purification Kit (Promega Biotech) and labeled by the DIG DNA Labeling and Detection Kit (Boehringer Mannheim). The digoxigenin-labeled probe was used to screen an oligo(dT)-primed λZAP cDNA library from salivary glands of *C. tentans* following the manufacturer's instructions. Positive clones were obtained and purified for sequence analysis.

DNA sequencing and sequence analysis

The inserted cDNA fragments of positive clones were amplified by PCR. The purified DNA fragments were used for sequencing with walking primers. The Taq DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems) was applied for the sequence reaction, and the sequencing gel was run on a 373A Automated DNA Sequencer (Applied Biosystems). The DNA sequences were analyzed by the University of Wisconsin Genetics Computer Group Sequence Analysis Programs.

For the evolutionary analysis, sequences were collected from the UniProt database, and homologs were identified by the BLAST algorithm. Alignment of the complete sequences was made with POA (Grasso and Lee 2004). Multiple alignments of WW and FF domains were constructed by aligning the sequences to the Pfam hidden Markov models (HMMs) PF00397 (WW domain) and PF01846 (FF domain); the domains were manually derived based on the Pfam definitions. Evolutionary trees were built with the Neighbor-joining method (Saitou and Nei 1987) using uncorrected distances and were rooted using P33203 (*Saccharomyces cerevisiae*). Bootstrap support for all nodes defining the ancestor of a domain exceeded 65%.

Results

Generation of mAb 2G8

Tissue culture cells of *C. tentans* were used to prepare nuclear protein extracts for producing monoclonal antibodies. The culture cells were broken by homogenization in a Nonidet P40 solution, the nuclei were isolated, and RNP extracts I and II were prepared as described in Materials and methods. In order to isolate RNA-binding proteins, the RNP extracts were mixed with ssDNA-agarose (Pinol-Roma et al. 1988). Loosely bound proteins were released with heparin, and the remaining proteins were eluted with a high-salt solution and used for mAb production. Spleen cells of immunized mice were fused with mouse myeloma Sp20 cells, and hybridoma clones were selected with a standard HAT procedure (Harlow and Lane 1988). Positive clones were identified by screening the hybridoma supernatants against the *C. tentans* RNP extract I in immunoblot experiments (Wurtz et al. 1996). Among the positive supernatants recognizing specific proteins, one of them, 2G8, bound to a 130 kDa protein (Fig. 1, lane 1; cf. total proteins in the nuclear extract in lane 5). We called this protein Ct-hrp130, or hrp130 for short.

hrp130 is a puff-specific, RNA-associated protein

In immunostaining experiments with isolated polytene chromosomes, we found that hrp130 is located preferentially in the BR3 puff. The salivary glands were pre-fixed in formaldehyde in the cold, and polytene chromosomes were isolated by a pipetting procedure. The chromosomes were incubated with the 2G8 hybridoma supernatant, and then challenged with an anti-mouse secondary antibody coupled to colloidal gold. The immuno-signal was silver enhanced. On chromosome IV, 2G8 reacted strongly with BR3; neither BR1, BR2 nor any other puffs gave a strong signal (Fig. 2a). Extensive analysis of the other three chromosomes occasionally revealed two additional, strongly labeled puffs, one on chromosome I (Fig. 2b) and one on chromosome II (Fig. 2c). Thus, hrp130 is

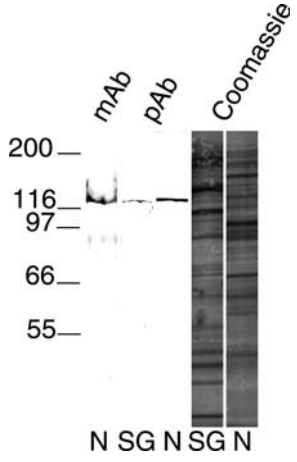

Fig. 1 Immunoblot analysis of hrp130 in *Chironomus tentans* nuclear extracts from tissue culture cells (*N*) and total salivary gland extracts (*SG*). The membrane was probed with either a monoclonal (*mAb*) (*lane 1*) or a purified polyclonal (*pAb*) antibody (*lanes 2, 3*) against hrp130. The positions of the molecular size markers (kilodaltons) are presented to the *left* and the Coomassie Blue stained extracts to the *right* (*lanes 4, 5*)

heavily accumulated in a low number of puffs, BR3 being the only puff consistently containing the protein in large amounts.

In order to reveal whether hrp130 is associated with RNA, isolated polytene chromosomes were RNase A treated prior to immunolabeling. The hrp130 protein could not be detected in BR3 after the RNase treatment (Fig. 3a vs c). It was also noted that BR3 regressed along with the disappearance of hrp130 (and RNP). Thus, hrp130 seems to be associated with RNA rather than DNA or chromatin.

By immunostaining RNase-treated chromosomes with an anti-RNA polymerase II antibody, we tested whether RNase treatment also releases RNA polymerases. The BRs were stained after the RNase treatment (Fig. 3b vs d), suggesting that hrp130 is bound to the growing RNAs rather than to the RNA polymerases.

In conclusion, hrp130 is heavily accumulated in BR3 on chromosome IV and occasionally in two additional chromosome puffs, one on chromosome I and one on chromosome II. The hrp130 protein is associated with RNA and is thus likely to be coupled to nascent RNA.

hrp130 is associated with nascent BR3 hnRNP particles

In order to study further the appearance of hrp130 in BR3, immunoelectron microscopy was applied. Salivary glands were pre-fixed and chromosome IVs were isolated. The chromosomes were incubated with 2G8 hybridoma supernatant, and challenged with an anti-mouse secondary antibody coupled to colloidal gold. The specimens were post-fixed, dehydrated, embedded in Agar resin, and examined and photographed in a transmission electron microscope.

The immunoelectron microscopy analysis was focused on the giant puff BR3. The active genes appeared as chromosomal loops (broad open arrows in Fig. 4a). In favorable projections, evenly distributed granular structures were seen along the loops (arrows in Fig. 4b). The granules are likely to correspond to nascent RNP complexes connected to spliceosomes (see electron tomographic reconstructions of the RNP complexes in Wetterberg et al. 2001). The tightly packed chromatin can be visualized as large compact areas (Fig. 4a,c), and a few connections between active genes and the dense chromatin

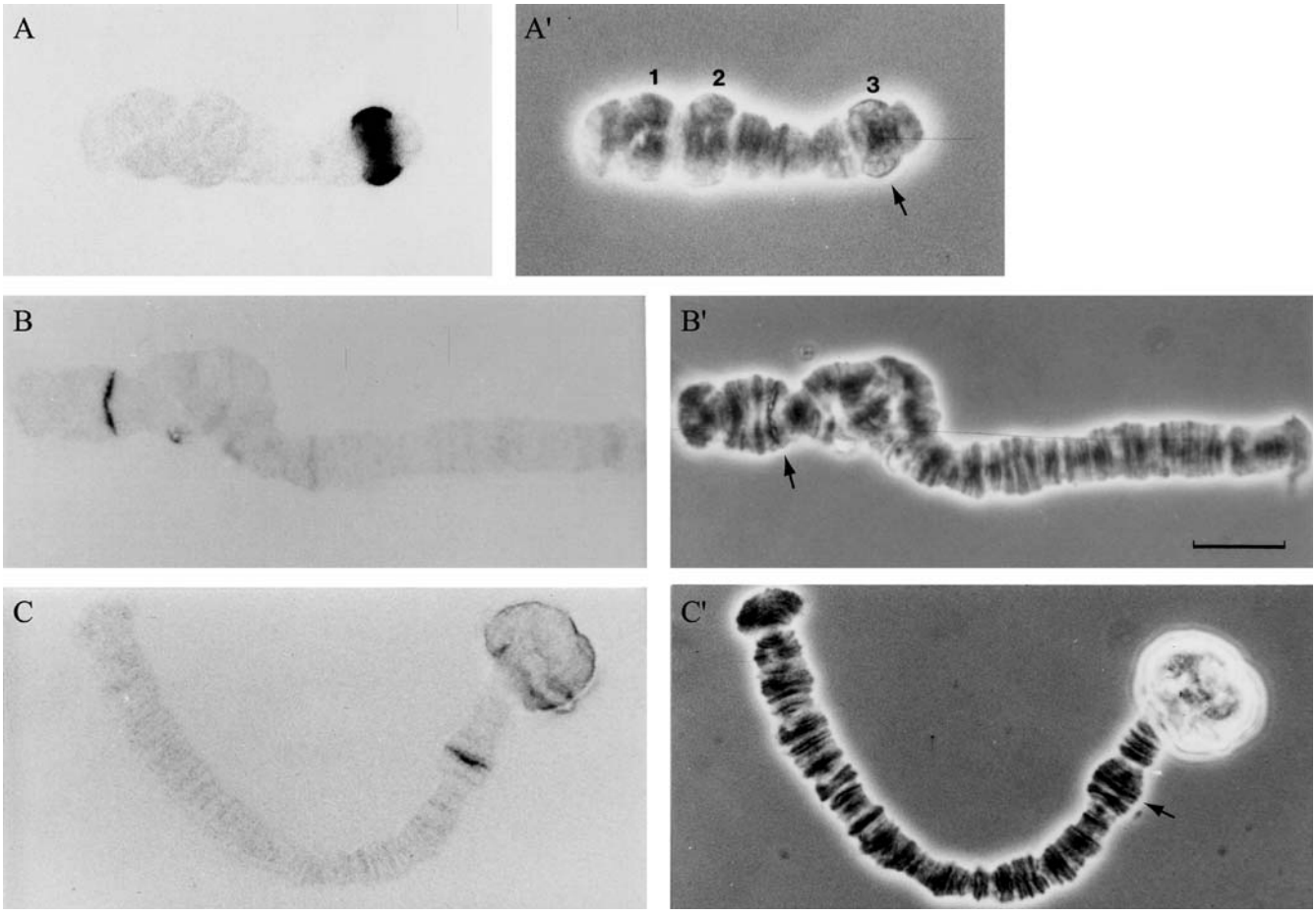

Fig. 2a–c Immunocytological localization of hrp130 on isolated polytene chromosomes. Salivary glands of *Chironomus tentans* were pre-fixed, and chromosomes I, II, and IV were isolated. The specimens were challenged with 2G8, and subsequently with a gold-conjugated secondary antibody. The immunostaining was visualized by silver enhancement. **a**,**a′** Chromosome IV; **b**,**b′** chromosome I; **c**, **c′** chromosome II. **a**–**c** Bright field; **a′**–**c′** phase contrast. The *numbers* in **a′** show the position of the three giant puffs, BR1, BR2, and BR3. The *arrows* in **a′**–**c′** indicate the position of the hrp130-positive loci. *Bar* represents 10 μm

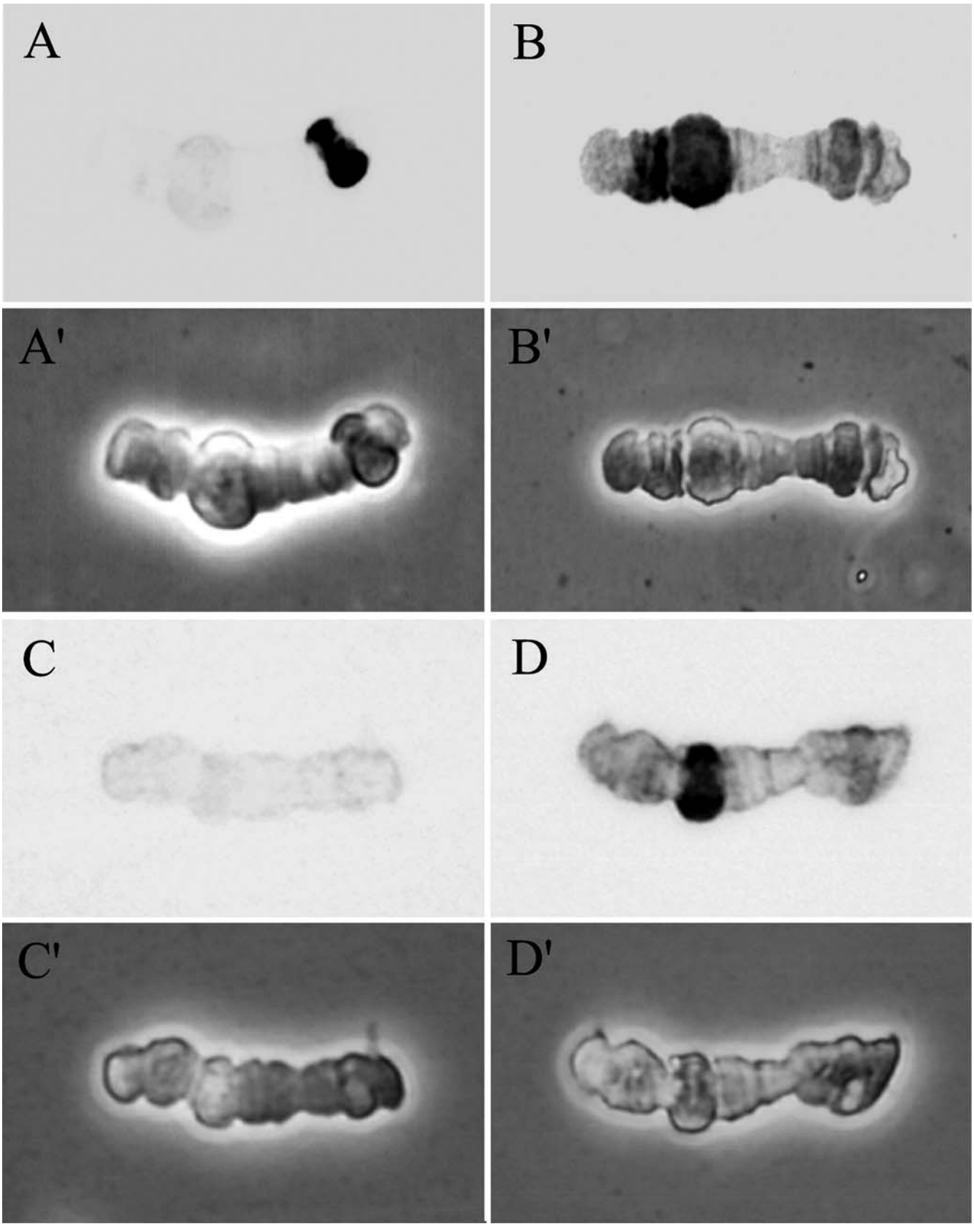

Fig. 3a–d Comparison of the immunocytological localization of hrp130 (**a**,**c**) and RNA polymerase II (**b**,**d**) on isolated chromosome IVs before and after RNase A treatment. Salivary glands of *Chironomus tentans* were pre-fixed and chromosome IVs were isolated. Half of them were treated with RNase A and subsequently all chromosomes were incubated with either 2G8 or an anti-Drosophila RNA polymerase IIo antibody. The immunostaining was visualized by the appropriate gold-conjugated antibody together with ensuing silver enhancement. **a**,**b** Without RNase A; **c**,**d** with RNase A treatment; **a′**–**d′** the corresponding chromosomes in phase contrast

have been marked in Fig. 4c (arrowheads). The impression at low magnification is that the inactive chromatin regions are dark and smooth, while the active genes are loosely packed and granular. The immunolabeling is clearly confined to the active genes (Fig. 4c) and appears along the genes (Fig. 4b). The junctions between compacted chromatin and active genes (arrowheads in Fig. 4c) did not appear labeled and are likely to correspond to partially unfolded chromatin upstream or downstream of the active genes. The BR1 and BR2 puffs on chromosome IV served as internal negative controls (Fig. 4d). The growing RNP particles in the large BRs are conspicuous; a few of them have been indicated by arrows in Fig. 4d. No labeling could be detected on the BR1 and BR2 RNP particles. We conclude that the immunoelectron microscopy experiments confirmed that hrp130 is essentially confined to

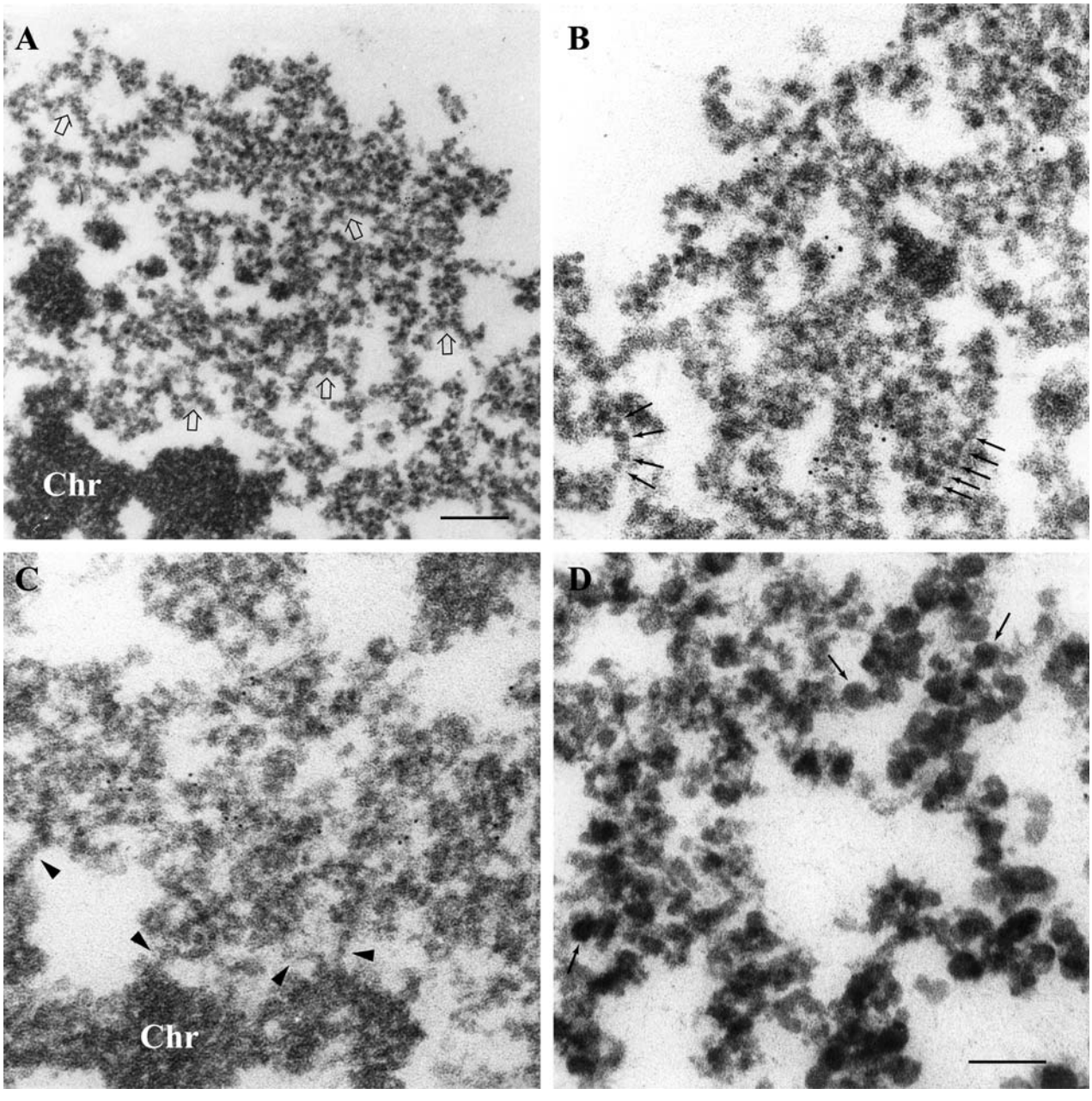

Fig. 4a–d Immunoelectron microscopic analysis of the distribution of hrp130 in the BR3 puff on chromosome IV. Salivary glands were pre-fixed and homogenized, and individual chromosome IVs were isolated manually. The 2G8 antibody was added, and later on a gold-conjugated secondary antibody. The specimens were then post-fixed, dehydrated, embedded and sectioned. **a**,**b** BR3. Transcription takes place along loops (*broad open arrows* in **a**). The individual transcription complexes can be seen as dense particles along the loops (*arrows* in **b**). Immunolabeling is recorded in regions with transcription loops, while the dense chromatin (*Chr*) and the junction regions between transcription loops and chromatin (*arrowheads* in **c**) are not labeled (**b**,**c**). The immunosignal over the active genes is more than 30 times stronger than that over background. **d** BR2. No labeling above background could be detected along the transcription loops with their large stalked RNP granules, some of them being indicated by *arrows*. *Bar* represents 200 nm in **a** and 100 nm in **d** (**b**–**d** the same magnification)

BR3 and, in addition, they demonstrate that hrp130 is present selectively along the transcriptionally active BR3 gene, further supporting the notion that hrp130 is a protein associated with nascent RNA.

hrp130 contains WW and FF domains

To obtain the cDNA sequence corresponding to hrp130, we screened two cDNA libraries. Both libraries were made from *C. tentans* salivary glands, but with different primers. First, the 2G8 antibody was used to screen a randomly primed λgt11 expression cDNA library. Four clones were isolated and sequenced, all having the same 392 bp insert. The insert was amplified by PCR and labeled with digoxigenin. The labeled DNA was used as a probe to screen an oligo(dT)-primed λZAP cDNA library. Two clones were collected, each with an insert of about 3.5 kb. Clones were sequenced and appeared to be identical. The complete hrp130 cDNA sequence was analyzed with sequence analysis programs and showed to encode a predicted 1028 amino acid protein. The predicted molecular mass of the protein is 121 kDa, which approximately corresponds to the estimate (130 kDa) obtained from gel electrophoresis.

To confirm that the predicted amino acid sequence corresponds to the 130 kDa protein initially identified through its interaction with the monoclonal antibody 2G8, a 100 amino acid C-terminal sequence of the predicted protein was expressed in bacteria. The peptide was purified and used to raise a polyclonal antibody, which was affinity-purified. In immunoblot analysis the polyclonal antibody showed a single band corresponding in size to the 130 kDa band detected using the 2G8 monoclonal antibody (Fig. 1, lanes 2, 3). Furthermore, when used on isolated chromosomes, the polyclonal antibody showed the same highly restricted binding pattern as the monoclonal antibody 2G8 (data not shown). Thus, we confirmed that the cloned sequence represents the hrp130 protein.

The amino acid sequence of hrp130 is presented in Fig. 5. There are three WW domains (WW1–3) in the N-terminal portion and six FF domains (FF1–6) in the C-terminal portion of the protein. WW domains are modules composed of about 40 amino acids (Sudol et al. 1995). The name alludes to two conserved tryptophan (W) residues, which are spaced 20–22 amino acids apart. The WW domains are found in many unrelated proteins, including signaling proteins, and bind to proteins with short linear proline-containing peptides (Chen and Sudol 1995; Zarrinpar and Lim 2000). The FF domains comprise about 50 amino acids and are characterized by two conserved phenylalanine (F) residues (Bedford and Leder 1999; Allen et al. 2002). The FF domains are also observed in many proteins and are likely to be involved in protein–protein interactions. Like in hrp130, the WW and FF domains often appear in the same protein (Bedford and Leder 1999); all these proteins seem to play a role in RNA metabolism.

```
1    MESDEGTSVV DNGEQEENVF SNDFTADFEN GWNNENADNQ PPMNPQQQNP
51   VQTENRFVPN NSNRGNNRGK GHRGGRGAGP NNWNPHQSGP PPQQPPQQIQ
101  QELWVETKTG DNKSYFYHAV TRETTWNRPD GPNIKVMTQA EFEAYTRQQM
                    WW1
151  RPVEQQQRPD QIKDPKMTMM PPNLMEHNTH LPPPTQQMPP FMPPFNANVS
201  PFGMHAPPHF NAPFPQWQPP NDPARMFDNR IDPKILAKAA EWTEHRAPDG
                                                    WW2
251  RPYYFSSARG ESVWERPQAL RELDEARAAF MHQQPPMTSS QGSITFDSAG
301  NMVKPGALMN KPPIEVADPG EKDRKRKEEI EKAKQQPAKP QDKTRPISST
351  PIAGTPWCVV WTGDSRVFFY NPSSRTSVWE RPQDLVGRAD VDKAVAVIPD
                    WW3
401  QLKKDGNVKE EVQVSEKANN PGIKVESEKS SEEEEEDDEV PTKKSKVEEQ
451  IAVPVKIQNA PPIVEKKVDV VKDPAVEAEL KAAKERAQIP LEVRVKQFKE
501  MLKEKEVSAF STWEKELHKI VFDQRYLLLA SKERKQVFEK YVKDRAEDER
                  FF1
551  REKRLKAQKK RDEFKALMEE ANLHSRSNFS DFCSRYSREE RYKGIEKMRE
                                    FF2
601  RENLFNDFLS ELRRREKDEK HLKKEQIRKE FFDLLKSHSE IDRHSHWMDI
                                                  FF3
651  KKKLDQDPRY KAITDSIQRE DYFYEYIKML KEERKKEKSK KAKKSEKKEK
701  KKKSKDKDRH KNESNNSNND AENKSNDESV KAEKEPQEND EIKPTDMEID
751  DDDIKSEIEI DKSGSESEKD QEDGEHSGTD EDSETEKARK DKERQQRAEA
801  SIKEREKQVQ MKLAEHLRDR DKERQHHKHD EAIRNFGALL ADLVRNPDLT
                                                  FF4
851  WKEAKKLLKK DHRYESDLER DERERLFNDH INLLAKKKRD KFREMLDEIA
901  TMELTSPWKE IKRLIRDDPR YSKFGNSDRC EREFRDYIRD KTANAKSEFK
                  FF5
951  ELLQECKLIT HKSYDLYKEN HNHLKEIEDI LKNDKRYLVL EHMPRDRSDM
                                FF6
1001 ILYYFKNLKK KGFPTPITQT VNNNRRKK
```

Fig. 5 The complete amino acid sequence of Ct-hrp130. Three WW and six FF domains have been identified (*named* and *underlined* in the figure)

The N-terminal moiety of hrp130 contains several short proline-rich sequences: PPNMPQQ, PPPQQPPQQ, PPNL, PPPTQ, PPFMPPFN, PPHF, PPND, and PPMT. Proline-rich sequences precede the first WWP domain and are also located between the first and second WW domains. These proline-rich motifs resemble proline-rich motifs in other WW-containing proteins: a PGF motif, a PPPPY motif and a PPLP motif, which have been shown to serve as ligands for WW domains (Chen and Sudol 1995; Ermekova et al. 1997; Bedford et al. 1998). Therefore, the proline-rich motifs in hrp130 may also be responsible for interaction with WW domains.

As hrp130 is associated with RNA, we have looked for RNA-binding motifs in the hrp130 sequence (Mattaj 1993; Burd and Dreyfuss 1994). No classical RNA-binding domain (RBD; also designated RNA recognition motif, RRM) was identified. However, there is a putative RGG box, which is located close to the N-terminus of the protein between amino acid residues 64 and 79. The sequence, RGNNRGKGHRGGRGAG, resembles an RGG box as it contains an RGG triplet and is rich in arginines and glycines.

In conclusion, hrp130 exhibits a characteristic structure with three WW domains in the N-terminal half and six FF domains in the C-terminal half of the protein. There are also short proline-rich regions in the N-terminal portion of

the protein. No classical RBD (RRM) is present, but there is a putative RRG box.

hrp130 is an evolutionarily conserved protein

Homology searches in the DNA and protein databases revealed that there are orthologs to hrp130 in *Anopheles gambiae* (Q7PMV0), *Drosophila melanogaster* (Q9VI14+; see legend to Fig. 6), *Caenorhabditis elegans* (Q95PX7) and *Homo sapiens* (O14776; known as CA150) (Fig. 6) . These four proteins are strikingly similar to hrp130, are of about the same size (914–1160 amino acids), and exhibit the same general organization with three WW domains followed by six FF domains.

The overall sequence similarity between the five proteins is in the interval 36–54%. However, the similarities are considerably higher when the individual WW and FF domains are compared. For example, when the WW3 domain of hrp130 in *C. tentans* is matched with the corresponding domain of the other four species, the similarities are within the range of 77–92% (a 26 amino acid long core sequence compared). Likewise, the FF1 domain of the *C. tentans* sequence is strikingly similar to the FF1 domains of the other four species, the highest similarity being 88% (*D. melanogaster*) and the lowest 61% (*C. elegans*) (a 49 amino acid long sequence compared).

To elucidate the relationships better, evolutionary trees for both the WW and FF domains were built with the neighbor-joining method using uncorrected observed evolutionary distances (Saitou and Nei 1987; see further Materials and methods). The WW tree is presented in Fig. 7a and the FF tree in Fig. 7b. The trees were rooted in a yeast homologous sequence (P33203; Prp40). The most striking result is that as a rule each individual WW domain (1, 2, or 3) and FF domain (1, 2, 3, 4, 5, or 6) is more similar to the corresponding domain of any of the other four species than it is to other domains of the same type in the same protein. In the trees, this is seen as a clustering of each individual WW domain (Fig. 7a) and of each individual FF domain (Fig. 7b). For example, when the WW1, WW2, and WW3 domains of hrp130 in *C. tentans* are compared with the corresponding WW domains in the *D. melanogaster* homolog, the similarity is 77, 81, and 92%, respectively. However, when, e.g., the WW1 domain in hrp130 is compared with the WW2 and WW3 domains in the same protein, the similarity is 31% and 38%, respectively.

The observation that each domain seems evolutionarily well conserved is strictly valid for four of the five species compared: the three insects (*C. tentans*, *D. melanogaster*, and *A. gambiae*) and *C. elegans*. Regarding the human sequence, the six FF domains and the WW3 domain also adhere to the general rule, while the human WW1 and WW2 domains have further diverged and form a cluster of their own. However, as the five proteins compared exhibit essentially the same number and organization of the WW and FF domains and a high sequence similarity, they are likely to be orthologs.

There are also proteins homologous to hrp130 in *Arabidopsis thaliana* (Q9LPD8) and *S. cerevisiae* (P33203; known as Prp40) (Fig. 6), but these contain two instead of three WW domains and four instead of six FF domains. Furthermore, the domains of these two proteins have diverged to such an extent that the individual WW and FF domains cannot be related to the corresponding domains of the five proteins considered above. Thus, it is not possible to conclude that the *Arabidopsis* and yeast homologs are orthologs. It should be pointed out that the yeast homolog, Prp40, has been thoroughly analyzed and is a splicing factor (see further Discussion).

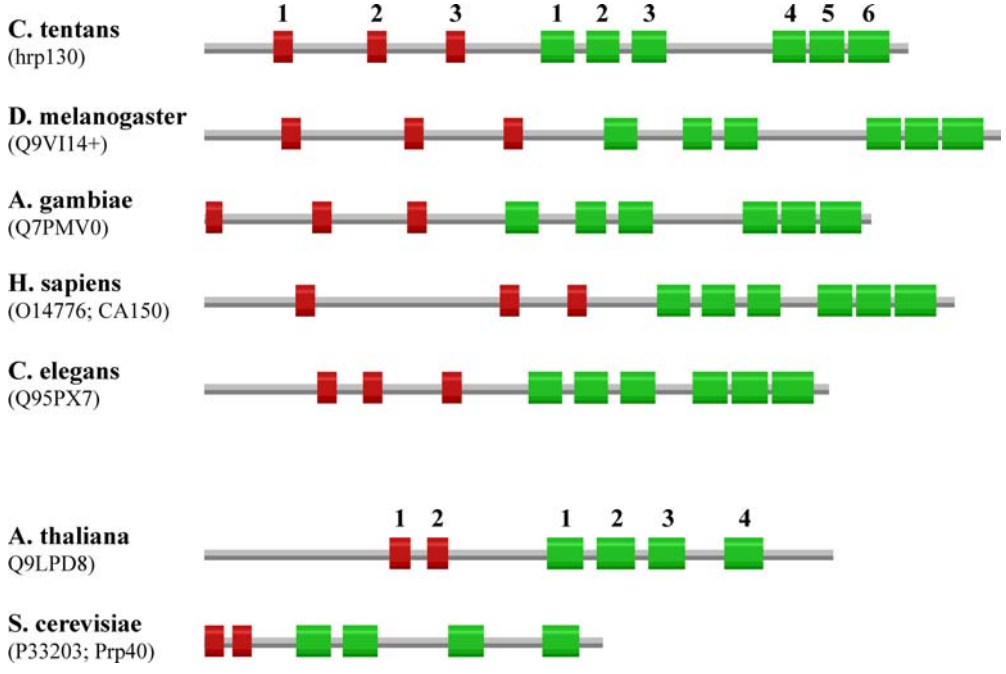

Fig. 6 Homologs to *Chironomus tentans* hrp130. The *C. tentans*, *Drosophila melanogaster*, *Anopheles gambiae*, *Caenorhabditis elegans*, and *Homo sapiens* proteins are highly homologous, and all contain three WW domains and six FF domains. The *Arabidopsis thaliana* and *Saccharomyces cerevisiae* proteins contain only two WW and four FFG domains and are less homologous to Ct-hrp130. The protein Accession numbers are given. The *D. melanogaster* sequence is denoted Q9VI14+, indicating that it corresponds to the previous entry of Q9VI14, which was recently split into two genes, Q9VI14 and Q95U01. We do not believe, however, that this split is warranted given both the intactness of the orthologs and the genomic sequence

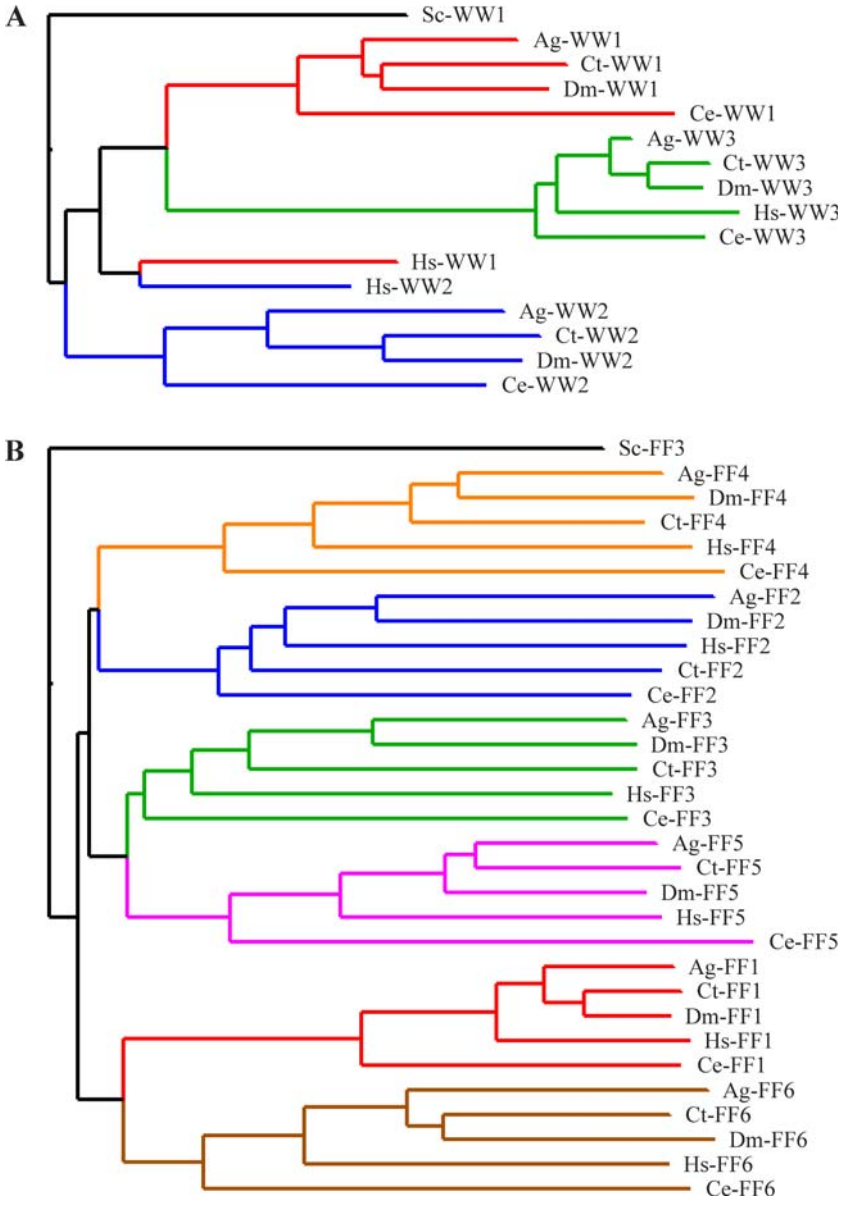

Fig. 7a,b Phylogenetic trees of the WW and FF domains in hrp130-like proteins from *Chironomus tentans*, *Drosophila melanogaster*, *Anopheles gambiae*, *Caenorhabditis elegans* and *Homo sapiens*. The WW domains are compared in **a** and the FF domains in **b**. The individual WW and FF domains of the various proteins are strongly clustered in the trees, indicating that the ancestral gene appeared early during evolution and has been maintained from *C. elegans* to human. *AgA. gambiae*, *CeC. elegans*, *CtC. tentans*, *DmD. melanogaster*, *HsH. sapiens*. For domain definitions, see Fig. 6

Discussion

hrp130 is heavily accumulated in BR3 and a few additional minor puffs

The hrp130 protein was consistently present in BR3, and occasionally also in two minor puffs, one on chromosome I and one on chromosome II. It has not been established whether the two minor puffs appear or vanish, or alternatively whether the composition of the proteins associated with the puff RNA varies. The striking accumulation of hrp130 in a few chromosomal sites is different from the broad distribution of three major RNA-binding proteins studied earlier: hrp36 (Visa et al. 1996), hrp45 (Alzhanova-Ericsson et al. 1996) and hrp23 (Sun et al. 1998). These three latter hnRNP proteins proved abundant in all three BRs as well as in a number of minor puffs on all four chromosomes. Although the distributions of these proteins were not identical (Wurtz et al. 1996), none of the proteins exhibited the striking puff specificity that has now been found for hrp130. The hrp130 protein rather behaves as the few minor hnRNP protein species in amphibian oocytes, which only appear in a restricted set of lampbrush loops (Sommerville et al. 1978) or are confined to one or both sets of giant loops (Roth and Gall 1987). In summary, the puff-specific distribution of hrp130 is strikingly different from that of most hnRNP proteins, which are bound to a large number of puffs on all chromosomes.

hrp130 is associated with nascent RNA

hrp130, together with other hnRNP proteins, was isolated from nuclear RNP extracts by ssDNA chromatography, which is an efficient method for heavily enriching for

RNA-binding proteins (Pinol-Roma et al. 1988). As nuclear RNA-binding proteins in general are adsorbed to the ssDNA, immunocytology experiments can help in deciding whether a specific protein is an hnRNP protein, a ribosomal protein, or a chromatin protein. In the case of hrp130, the protein was present in one or a few chromosomal puffs, especially in BR3. As the BR3 gene encodes a secretory protein (Dignam and Case 1990; Paulsson et al. 1990), hrp130 is likely to be associated with pre-mRNA. The hrp130 protein was released from BR3 by RNase treatment, suggesting that it is indeed linked to the RNA in the puff. Such a conclusion was further corroborated by the immunoelectron microscopy data. The immunolabeling was detected exclusively along the BR3 transcription loops; adjacent chromatin regions were unlabeled. Thus, hrp130 is associated with nascent pre-mRNA in BR3.

hrp130 is a homolog of the human protein CA150

The *C. tentans* hrp130 protein is evolutionarily conserved, and the same general sequence organization, i.e., three WW domains followed by six FF domains, is found in hrp130 homologs identified in *D. melanogaster*, *A. gambiae*, *C. elegans* and *H. sapiens*. The conservation of the sequence is considerable between *C. elegans* and human; the corresponding WW domains show a sequence similarity of 80–90% while the similarity between the corresponding FF domains amounts to 60–90%. However, the WW as well as FF domains present in the same protein are considerably more diverged (similarity 30–40% and 20–25%, respectively). It is thus remarkable that the divergence between the individual WW domains as well as between the individual FF domains has been maintained to such a high extent from *C. elegans* to human. Thus, the organization of the hrp130-like proteins into three WW and six F domains was established early during evolution, and the various domains were defined before *C. elegans* and insects diverged. The fact that the various domains have been conserved during evolution to such a high degree suggests that each individual domain plays an important and specific physiological role in the protein. Only the human protein, CA150, has been functionally investigated, and it is known to be a transcriptional repressor (see below).

hrp130—a putative repressor of transcription elongation

The striking conservation of the hrp130/CA150 protein during evolution suggests that *C. tentans* hrp130 exerts a role similar to that of CA150. It was early established that CA150 affects the elongation stage during transcription (Sune et al. 1997). The observation that hrp130 is distributed along the active BR3 genes is in agreement with such a role. More recently, it has been shown that CA150 represses elongation (Sune and Garcia-Blanco 1999; M. Garcia-Blanco, pers. commun.), which can a priori be accomplished by affecting the RNA polymerase itself or by displacing/suppressing positive transcription elongation factors (P-TEFs) or recruiting/activating negative transcription elongation factors (N-TEFs) (for discussion, see Goldstrohm et al. 2001). It has been shown that CA150 interacts, via its FF domains, directly and specifically with the phosphorylated C-terminal domain (phospho-CTD) of the polymerase (Carty et al. 2000). Furthermore, CA150 can associate with several P-TEFs (e.g., P-TEF-b, Tat-SF1, and TFIIF) and at least one putative N-TEF, viz. Splicing factor 1 (SF1) (Goldstrohm et al. 2001). An interaction with SF1 is particularly interesting as the RNA-binding SF1 could connect CA150 to the nascent RNA. Thus, the CA150–SF1 complex could bind to the elongating transcription complex, recognizing both the phospho-CTD of the polymerase and the growing transcript. Such a dual interaction would be analogous to the mode of binding of other elongation factors repressing transcription (e.g., the Nrd–Nab and DSIF–NELF complexes; for discussion and references, see Goldstrohm et al. 2001). Our observation that the presence of hrp130 on the polytene chromosomes is sensitive to RNase treatment would be in agreement with the possibility that hrp130/CA150 is part of a complex that bridges the polymerase and the nascent transcript. The stability of the complex would then be dependent on an intact RNA molecule. We conclude that hrp130/CA150 is likely to act as a repressor of elongation, being in contact with both the RNA polymerase and the growing transcript.

A role in adjusting elongation rate to splicing?

The fact that CA150 can bind to SF1, best known as an essential, branch-point binding splicing factor (Krämer and Utans 1991), suggests that CA150 could also be functionally coupled to RNA splicing. Such a hypothesis is further supported by the observation that CA150 has been recorded in spliceosomes (Neubauer et al. 1998) and been found associated with small nuclear RNPs (M. Garcia-Blanco, pers. commun.). Moreover, it should be recalled that CA150 exhibits a striking structural similarity to the yeast splicing factor Prp40 with its two WW and four FF domains (Fig. 6). It is noteworthy that PRP40 binds to the phospho-CTD of RNA polymerase II (Morris and Greenleaf 2000) as well as to U1snRNP (Kao and Siliciano 1996; Neubauer et al. 1997) and the branch-point binding Msl5 protein (Abovich and Rosbash 1997). Thus, the elongation repressor CA150 is likely to be connected to the splicing machinery in some way.

In this context, it is interesting to consider our finding that hrp130 is heavily accumulated in BR3, known to be exceptionally active in splicing. The BR3 gene contains 38 introns distributed evenly along the 10.9 kb gene (Paulsson et al. 1990). The template harbors 20–25 transcription complexes, most or all with a spliceosome, i.e., essentially all growing transcripts on the BR3 gene are involved in splicing at a given time point (Wetterberg et al.

2001). This is drastically different from the situation on the BR1 and BR2 genes, which are also very active in transcription. The BR1 and BR2 genes are 3–4 times longer (35–40 kb) than the BR3 gene but only contain four introns (Wieslander 1994). Furthermore, on the BR1 and BR2 genes spliceosomes only appear on the most promoter-proximal and promoter-distal portions, removing the three 5′ end introns and the single 3′ end intron, respectively; the long central portion of the gene corresponding to the 30 kb exon 4 is to a large extent devoid of spliceosomes (Kiseleva et al. 1994). Thus, the splicing intensity in BR3 is exceptional and virtually every transcription complex along the intensely transcribed BR3 gene carries a spliceosome.

The in vivo kinetics of the removal of the 38 introns from the BR3 transcript has been carefully analyzed (Wetterberg et al. 1996). Essentially all introns are removed co-transcriptionally. Furthermore, as a rule, the introns first transcribed are also the ones first removed. It was early suggested that the transcription process itself helps to accomplish correct splicing as the splice sites become available for splicing in the proper order (Aebi and Weissmann 1987). Such a mechanism would be especially helpful if the gene contains many introns as in the case of the BR3 gene. Furthermore, to be optimal the mechanism would require that the elongation rate is modulated to make sure that the splicing process keeps pace with the appearance of new, but not too many, splice sites. If for some reason splicing is slowed down, transcription elongation should be accordingly repressed.

It is an interesting possibility that the CA150–SF1 complex is the crucial factor adjusting the transcription rate to the splicing process. Such regulation could again be particularly pertinent to genes like the BR3 gene with many introns and a high level of transcription, i.e., a high density of spliceosomes bound to transcripts along the gene. Due to the exceptional need for splicing components at such a gene, it is conceivable that there would be a shortage of splicing factors, resulting in a reduced splicing efficiency. The polyteny in the BR3 case could further aggravate the situation; there are about 10,000 active BR3 genes in the puff (Daneholt and Edström 1967). If hrp130/CA150 were then recruited, the elongation rate would be reduced and the accuracy of splicing secured.

hrp130/CA150 is strongly evolutionarily conserved, which suggests that it is a transcription factor important for the function of many genes. The fact that CA150 is present in all tissues investigated (Bohne et al. 2000) also indicates a more general function. Still hrp130/CA150 seems to be confined to just a few chromosomal sites in the salivary glands of *C. tentans*. One possible explanation is that hrp130/CA150 is ready to be recruited to active RNA polymerases whenever the elongation rate of RNA polymerase II has to be reduced to optimize gene expression, the adjustment of transcription to splicing in BR3 being one example.

Acknowledgements We thank Lars Wieslander for providing us with *Chironomus tentans* cDNA libraries. The anti-RNA polymerase II antibody was a gift from Arno L. Greenleaf (Duke University Medical Center, Durham, N.C.). We are grateful to Sergej Masich for computer support and to Birgitta Björkroth and Lise-Marie Fjelkestam for technical assistance. This study was supported by the Swedish Research Council, Human Frontier Science Program Organization, Knut and Alice Wallenberg Foundation, Kjell and Märta Beijer Foundation, Ingabritt and Arne Lundberg Foundation and the Karolinska Institute Research Funds.

References

Abovich N, Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89:403–412

Aebi M, Weissman C (1987) Precision and orderliness in splicing. Trends Genet 3:102–107

Allen M, Friedler A, Schon O, Bycroft M (2002) The structure of an FF domain from human HYPA/FB11. J Mol Biol 323:411–416

Alzhanova-Ericsson AT, Sun X, Visa N, Kiseleva E, Wurtz T, Daneholt B (1996) A protein of the SR family of splicing factors binds extensively to exonic Balbiani ring pre-mRNA and accompanies the RNA from the gene to the nuclear pore. Genes Dev 10:2881–2893

Baurén G, Belikov S, Wieslander L (1998) Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3′-end formation and excision of the 3′-terminal intron. Genes Dev 12:2759–2769

Bedford MT, Leder P (1999) The FF domain: a novel motif that often accompanies WW domains. Trends Biochem Sci 24:264–265

Beford MT, Chan DC, Leder P (1997) FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. EMBO J 16:2376–2383

Bentley D (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. Curr Opin Cell Biol 11:347–351

Björkroth B, Ericsson C, Lamb MM, Daneholt B (1988) Structure of the chromatin axis during transcription. Chromosoma 96:333–340

Bohne J, Cole SE, Sune C, Lindman BR, Ko VD, Vogt TF, Garcia-Blanco MA (2000) Expression analysis and mapping of the mouse and human transcriptional regulator CA150. Mamm Genome 11:930–933

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science 265:615–621

Carty SM, Goldstrohm AC, Sune C, Garcia-Blanco MA, Greenleaf A (2000) Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. Proc Natl Acad Sci USA 97:9015–9020

Chen HI, Sudol M (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc Natl Acad Sci USA 92:7819–7823

Daneholt B (2001) Assembly and transport of a premessenger RNP particle. Proc Natl Acad Sci USA 98:7012–7017

Daneholt B, Edström J-E (1967) The content of deoxyribonucleic acid in individual polytene chromosomes of *Chironomus tentans*. Cytogenetics 6:350–356

Dignam SS, Case ST (1990) Balbiani ring 3 in *Chironomus tentans* encodes a 185-kDa secretory protein which is synthesized throughout the fourth larval instar. Gene 88:133–140

Dreyfuss G, Matunis MJ, Pinol Roma S, Burd CG (1993) hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem 62:289–321

Ermekova KS, Zambrano N, Linn H, Minopoli G, Gertler F, Russo T, Sudol M (1998) The WW domain of neural protein FE65 interacts with proline-rich motifs in mena, the mammalian homolog of *Drosophlia* enabled. J Biol Chem 272:32869–32877

Fakan S (1994) Perichromatin fibrils are in situ forms of nascent transcript. Trends Cell Biol 4:86–90

Goldstrohm AC, Albrecht TR, Sune C, Bedford MT, Garcia-Blanco MA (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. Mol Cell Biol 21:7617–7628

Grasso C, Lee C (2004) Combining partial order alignment and progressive multiple sequence alignment increases alignment speed and scalability to very large alignment problems. Bioinformatics 20:1546–1556

Harlow E, Lane D (1988) Antibodies, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Hirose Y, Manley JL (2000) RNA polymerase II and the integration of nuclear events. Genes Dev 14:1415–1429

Kabisch R, Bautz EKF (1983) Differential distribution of RNA polymerase B and nonhistone chromosomal proteins in polytene chromosomes of *Drosophila melanogaster*. EMBO J 2:395–402

Kao HY, Siliciano PG (1996) Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol 16:960–967

Kiseleva E, Wurtz T, Visa N, Daneholt B (1994) Assembly and disassembly of spliceosomes along a specific pre-messenger RNP fiber. EMBO J 13:6052–6061

Krämer A, Utans U (1991) Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. EMBO J 10:1503–1509

Krecic AM, Swanson MS (1999) hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol 11:363–371

Lacroix JC, Azzouz R, Boucher D, Abbadie C, Pyne CK, Charlemagne J (1985) Monoclonal antibodies to lampbrush chromosome antigens of *Pleurodeles waltlii*. Chromosoma 92:69–80

Lei EP, Krebber H, Silver PA (2001) Messenger RNAs are recruited for nuclear export during transcription. Genes Dev 15:1771–1782

Lezzi M, Meyer B, Mähr R (1981) Heat shock phenomena in *Chironomus tentans* I. In vivo effects of heat, overheat, and quenching on salivary chromosome puffing. Chromosoma 83:327–339

Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. Nature 416:499–506

Mattaj I (1993) RNA recognition: a family matter? Cell 73:837–840

Matunis EL, Matunis MJ, Dreyfuss G (1993) Association of individual hnRNP proteins and snRNPs with nascent transcripts. J Cell Biol 121:219–228

Morris DP, Greenleaf AL (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem 275:39935–39943

Morris DP, Lee JM, Sterner DE, Brickey WJ, Greenleaf AL (1997) Assaying CTD kinases in vitro and phosphorylation-modulated properties of RNA polymerase II in vivo. Methods Companion Methods Enzymol 12:264–275

Neubauer G, Gottschalk A, Fabrizio P, Seraphin B, Lührmann R, Mann M (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc Natl Acad Sci USA 94:385–390

Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuk P, Sleeman J, Lamond A, Mann M (1998) Mass spectrometry and EST-database searching allows characterization of the multiprotein spliceosome complex. Nat Genet 20:46–50

Paulsson G, Lendahl U, Galli J, Ericsson C, Wieslander L (1990) The Balbiani ring 3 gene in *Chironomus tentans* has a diverged repetitive structure split by many introns. J Mol Biol 211:331–349

Pelling C (1964) Ribonukleinsäure-Synthese der Riesenchromosomen. Autoradiographische Untersuchungen an *Chironomus tentans*. Chromosoma 15:71–122

Pinol-Roma S, Choi YD, Matunis MJ, Dreyfuss G (1988) Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev 2:215–227

Pinol-Roma S, Swanson MS, Gall JG, Dreyfuss G (1989) A novel heterogeneous nuclear RNP protein with a unique distribution on nascent transcripts. J Cell Biol 109:2575–2587

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. Cell 108:501–512

Roth MB, Gall JG (1987) Monoclonal antibodies that recognize transcription unit proteins on newt lampbrush chromosomes. J Cell Biol 105:1047–1054

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sommerville J, Crichton C, Malcolm D (1978) Immunofluorescent localization of transcriptional activity on lampbrush chromosomes. Chromosoma 66:99–114

Strässer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E (2002) TREEX is a conserved complex coupling transcription with messenger RNA export. Nature 417:304–307

Sudol M, Chen HI, Bougeret C, Einbond A, Bork P (1995) Characterization of a novel protein-binding module—the WW domain. FEBS Lett 369:67–71

Sun X, Alzhanova-Ericsson AT, Visa N, Aissouni Y, Zhao J, Daneholt B (1998) The hrp23 protein in the Balbiani ring premRNA particles is released just before or at the binding the particles to the nuclear pore complex. J Cell Biol 142:1181–1193

Sune C, Garcia-Blanco MA (1999) Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. Mol Cell Biol 19:4719–4728

Sune C, Hayashi T, Liu Y, Lane W, Young RA, Garcia-Blanco MA (1997) CA150, a nuclear protein associated with RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. Mol Cell Biol 17:6029–6039

Visa N, Alzhanova-Ericsson AT, Sun X, Kiseleva E, Björkroth B, Wurtz T, Daneholt B (1996) A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. Cell 84:253–264

Wetterberg I, Baurén G, Wieslander L (1996) The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. RNA 2:641–651

Wetterberg I, Zhao J, Masich S, Wieslander L, Skoglund U (2001) In situ transcription and splicing in the Balbiani ring 3 gene. EMBO J 20:2564–2574

Wieslander L (1994) The Balbiani ring multigene family: coding repetitive sequences and evolution of a tissue-specific cell function. Prog Nucleic Acid Res Mol Biol 48:275–313

Wurtz T, Kiseleva E, Nacheva G, Alzhanova-Ericsson AT, Rosén, A, Daneholt B (1996) Identification of two RNA-binding proteins in Balbiani ring premessenger ribonucleoprotein granules and presence of these proteins in specific subsets of heterogeneous nuclear ribonucleoprotein particles. Mol Cell Biol 16:1425–1435

Wyss C (1982) Ecdysterone, insulin and fly extract needed for the proliferation of normal Drosophila cells in defined medium. Exp Cell Res 139:297–307

Zarrinpar A, Lim WA (2000) Converging on proline: the mechanism of WW domain peptide recognition. Nat Struct Biol 7:611–613