Isolation and characterization of the androgen-dependent mouse cysteine-rich secretory protein-1 (*CRISP-1*) gene

Uta SCHWIDETZKY, Wolf-Dieter SCHLEUNING and Bernard HAENDLER*
Research Laboratories of Schering AG, D-13342 Berlin, Germany

In mice, cysteine-rich secretory protein-1 (CRISP-1) is mainly found in the epididymis and also, to a lesser extent, in the salivary gland of males, where androgens control its expression. We have now isolated and characterized overlapping phage clones covering the entire length of the *CRISP-1* gene. DNA sequencing revealed that the gene is organized into eight exons, ranging between 55 and 748 bp in size, and seven introns. All exon–intron junctions conformed to the GT/AG rule established for eukaryotic genes. The intron length, as determined by PCR, varied between 1.05 and 4.0 kb so that the *CRISP-1* gene spans over 20 kb of the mouse genome. The transcription-initiation site was determined by primer extension and localized at the expected distance downstream of a consensus TATA box. Approximately 3.7 kb of the *CRISP-1* promoter region were isolated and sequenced, and several stretches fitting the androgen-responsive element consensus were found. Those that most resembled the consensus were analysed by electrophoretic mobility-shift assay and found to form specific complexes with the liganded androgen receptor *in vitro*, but with different affinities. Putative binding elements for the transcription factors Oct, GATA, PEA3, CF1, AP-1 and AP-3 were also found in the promoter region.

INTRODUCTION

Murine cysteine-rich secretory protein-1 (CRISP-1) was originally isolated from the epididymis [1], where spermatozoa undergo final stages of maturation, including the acquisition of motility and of fertilizing competence [2,3] and the remethylation of spermatogenesis-specific genes [4]. It is probably identical with MEP7, an abundant protein made by the principal cells and taken up by the clear cells of the epididymis [5,6]. CRISP-1 exhibits a two-domain structure [1]. The N-terminal domain contains six pairwise-linked cysteines and the C-terminal one ten cysteines, with a spacing characteristic of the CRISP family. *CRISP-1* transcripts have been detected in the epididymis and, to a lesser extent, in the salivary gland of male mice [7,8].

The human and rat CRISP-1 counterparts display 41 and 70% respectively amino acid sequence identity with the mouse form and are likewise abundant in the epididymis [9–13]. The rat protein, which has several names including DE, binds to spermatozoa, possibly by interacting with specific receptors [14,15]. The existence of complementary sites for DE on the egg plasma membrane as well as the results of immunization studies indicate a role for this protein in the fertilization process [16–18], but no equivalent data are available for mouse or human.

Other CRISP family members include CRISP-2 (also called Tpx-1 in mouse, TPX1 in human and AA1 in guinea pig), which is expressed in male haploid germ cells and accumulates in the acrosome [19–21], and CRISP-3, which was originally identified as a mouse salivary-gland cDNA [7,8]. Recent studies indicate that murine CRISP-3 is also expressed in pre-B-cells [22] and human CRISP-3 in the specific granules of neutrophils [23], in the prostate, the ovary, the pancreas, the thymus and the colon [9]. No precise function has yet been demonstrated for CRISP-2 or CRISP-3 but a non-specific defence role has been proposed for CRISP-3, in analogy with the distantly related pathogenesis-related proteins of plants [24]. Another member of the CRISP family is helothermine, a toxin that causes hypothermia and was isolated from the salivary gland of the Mexican beaded lizard *Heloderma horridum* [25]. Recent studies indicate that helothermine can block cardiac and skeletal ryanodine-sensitive Ca^{2+} channels *in vitro* [26], a finding that may provide clues to the function of other family members.

The sexual dimorphism observed in the salivary gland and the high expression level determined in the epididymis suggest that the male sex hormone plays an important role in the expression of the murine *CRISP-1* gene. This was confirmed *in vivo* by analysing the salivary-gland RNAs of castrated male and androgen-treated female mice [7,8]. Expression of the murine *CRISP-3* gene is also positively regulated by androgens in the salivary gland, but to a lesser extent [7,8]. Indeed putative androgen-responsive elements (AREs) were found in the promoter of the *CRISP-3* gene, which has recently been isolated [27] and mapped to chromosome 17, in the vicinity of the *CRISP-1* and *CRISP-2/Tpx-1* genes [28].

In order to take the analysis of the elements underlying the tissue-specific expression and androgen control of the *CRISP* genes a step further, we isolated and characterized the mouse *CRISP-1* gene, including approximately 3.7 kb of promoter region. Putative AREs were identified and analysed by electrophoretic mobility-shift assay (EMSA) using heterologously expressed androgen receptor.

EXPERIMENTAL

Materials

Fresh epididymis was obtained from NMRI mice kept at the in-house facility. The C57 black mouse liver genomic library was from Clontech. Restriction endonucleases, the reverse transcriptase, RNase A and the 5′ DNA end-labelling kit were purchased from Boehringer-Mannheim. Poly[d(I-C)] was from

Abbreviations used: ARE, androgen-responsive element; dATP[αS], deoxyadenosine 5′-[α-thio]triphosphate; CRISP, cysteine-rich secretory protein; EMSA, electrophoretic mobility-shift assay; poly(A)$^+$, polyadenylated; DTT, dithiothreitol.

* To whom correspondence should be sent at the following address: Institute of Cellular and Molecular Biology, Schering AG, D-13342 Berlin, Germany.

Pharmacia. Pefabloc SC was from Biomol. *Taq* polymerase and the PCR kit were from Perkin–Elmer–Cetus. T7 gene 6 exonuclease and the Sequenase kit version 2.0 were obtained from USB. The ribonuclease inhibitor RNAsin was from Promega and the Fast Track mRNA Isolation kit from InVitrogen. The pBluescript plasmid was from Stratagene whereas the pAcSG His NT-C transfer vector, the BaculoGold DNA, the wild-type AcNPV virus and the Sf9 cells were from PharMingen. The prestained protein molecular-mass markers were from Bio-Rad. Oligodeoxynucleotides were made on an Applied Biosystems 391A DNA synthesizer. The labelled nucleotides [α-^{32}P]dCTP ($\sim$ 3000 Ci/mmol), [γ-^{32}P]ATP ($\sim$ 3000 Ci/mmol) and [^{35}S]-deoxyadenosine 5′-[α-thio]triphosphate (dATP[αS]) ($>$ 1000 Ci/mmol) were purchased from Amersham. R1881 was from Dupont–NEN. 5α-Dihydrotestosterone, aprotinin, pepstatin, the anti-rabbit IgG alkaline phosphate conjugate and the SigmaFast tablets were from Sigma. Agarose was from Life Technologies. Nonidet P40 was from LKB. The Biodyne A (1.2 μm) hybridization membranes were obtained from Pall and the BA85 membranes from Schleicher and Schuell. The X-OMAT and Biomax films used for autoradiography were from Kodak. The University of Wisconsin GCG package [29] was used for sequence analysis.

Isolation of phage clones

Between 0.5×10^6 and 1×10^6 phages from a mouse genomic DNA library established in the λEMBL3 vector were replica-plated on to nylon membranes and screened as described [27]. The initial probe was derived from a 322 bp-long *Nco*I fragment isolated from the 5′ end of the *CRISP-3* cDNA which shares 95% sequence identity with the 5′ end of *CRISP-1*, using [α-^{32}P]dCTP and *Taq* polymerase [27]. Additional screenings were performed with a probe derived from an internal 95 bp-long PCR fragment amplified from *CRISP-1* cDNA (positions 449–543 in [8]) and with the following oligonucleotides labelled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP: 5′-TCCTTGATAATTGCCAAC-3′, 5′-GTGCCAGTTGTCCTG-ATC-3′ and 5′-TGTGGACATGAAGATAAG-3′ (positions 542–559, 594–611 and 641–658 in the *CRISP-1* cDNA [8]). Phage plaques scoring positive were isolated and analysed by following the strategy described in [27].

DNA sequencing

The dideoxy chain-termination method [30] was used to sequence double-stranded plasmid DNA with a Sequenase kit and [^{35}S]dATP[αS]. Phage DNA was sequenced after preparing single-stranded template by T7 gene 6 exonuclease digestion or by asymmetric PCR [27]. The sequence of both DNA strands was determined.

Determination of intron size

Primer pairs matching two consecutive exons or their flanking regions were used to amplify the intervening regions by PCR [27]. For each intron, two different pairs were used for a more accurate determination. The size of the generated fragments was determined by comparison with a DNA molecular-mass standard and the value corrected with respect to the exact location of the primers used.

Determination of the transcription-initiation site

Polyadenylated [Poly(A)$^+$] RNA was isolated from mouse epididymis using the FastTrack mRNA Isolation kit. A labelled antisense oligonucleotide (5′-TTAGTGGTTGAAAGTTTCTC-AAGACGATTT-3′, positions 100–129 in the *CRISP-1* cDNA, [8]) was mixed with 2 μg of poly(A)$^+$ RNA and precipitated in ethanol. Primer extension and analysis of the generated products were performed as described [27].

Preparation of androgen-receptor-containing extracts from insect cells

The human androgen receptor cDNA [31] was introduced into the pAcSG His NT-C baculovirus transfer plasmid. Recombinant baculovirus stock was obtained by homologous recombination after cotransfection of Sf9 insect cells with the transfer vector and BaculoGold DNA, following the manufacturer's instructions. For the preparation of extracts, Sf9 cells were infected with recombinant baculovirus (3 plaque-forming units/cell) in the presence of the synthetic androgen R1881 (0.1 μM). For control experiments, wild-type AcNPV virus was used. After 48 h the cells were harvested, centrifuged for 15 min at 140 ***g*** and 4 °C, and the pellet was resuspended in 1/12th vol. of cold 20 mM Tris/HCl, pH 7.5, containing 0.5 mM EDTA, 400 mM KCl, 20 mM sodium tungstate, 2 mM dithiothreitol (DTT), 20% glycerol, 0.3 μM aprotinin, 1 μM pepstatin and 50 μM Pefabloc SC. The cells were lysed through three freeze–thaw cycles in liquid nitrogen and centrifuged for 1 h at 100000 ***g*** and 4 °C. The supernatant was shock-frozen in small aliquots and stored at −70 °C. For the supershift assays, the cell pellet was resuspended in 1/15th vol. of cold 50 mM Tris/HCl, pH 7.5, containing 0.1 mM EDTA, 400 mM KCl, 5 mM DTT, 20% glycerol, 0.5 mM Pefabloc SC and 0.1 μM 5α-dihydrotestosterone. After centrifugation for 45 min at 100000 ***g*** and 4 °C, the supernatant was recovered and precipitated by adding $(NH_4)_2SO_4$ (0.3 g/ml supernatant) and stirring for 30 min at 4 °C. The precipitated proteins were recovered by 20 min centrifugation at 50000 ***g***, resuspended in buffer D (1/75th of the original volume) and dialysed against the same buffer (buffer D is 20 mM Hepes, pH 7.9, containing 0.2 mM EDTA, 100 mM KCl, 2 mM DTT, 20% glycerol, 0.05% Nonidet P40 and 0.5 mM Pefabloc SC [32]) supplemented with 0.1 M 5α-dihydrotestosterone. Aliquots were shock-frozen. The protein concentration of the extracts was determined by the method of Bradford [33]. Western-blot analysis was carried out [34] using the antiserum ARP4 raised against amino acids 250–268 of the human androgen receptor [35]. A 1:1000 dilution of ARP4 and 1:8000 dilution of anti-rabbit IgG alkaline phosphate conjugate were used.

EMSA

The following oligonucleotides covering putative AREs of the *CRISP-1* promoter starting at positions −1270 and −1253 were synthesized: BS1, 5′-ATGCATTG AGTAGATCTGTTCTATGGTACATCTTGTTCA CATAG-ACA-3 and its complement BS2; BS5, 5′-ATGCATTG-GGTACATCTTGTTCACATAGACA-3′ and its complement BS6; BS9, 5′-ATGCATTGAGTAGATTCTGTTCTCATAG-ACA-3′ and its complement BS10 (the regions corresponding to the ARE consensus are underlined). Labelling was as above with T4 polynucleotide kinase and [γ-^{32}P]ATP. Complementary oligonucleotides were annealed after 10 min heating at 90 °C by cooling down to room temperature over 1 h. The binding reaction took place in a final volume of 10 μl in 20 mM Hepes, pH 7.9, containing 0.1 mM EDTA, 5 mM $MgCl_2$, 2 mM DTT, 20% glycerol and 0.25 mM Pefabloc SC. Human androgen receptor extract (1 μl of a 6 μg of protein/μl preparation) was incubated with R1881 (1 μM) for 15 min on ice. Then a mixture made of 1 μl of the annealed labelled oligonucleotides (50000 c.p.m.), 1 μl

of 1 μg/μl poly[d(I-C)] and 1 μl of 20% Ficoll was added. For competitive experiments, annealed unlabelled oligonucleotides were also included. After 1 h incubation at room temperature the reaction mixture was analysed on a 4% polyacrylamide/bisacrylamide (19:1)/5% glycerol/0.5 × TBE gel [1 × TBE is 90 mM Tris/HCl (pH 8)/25 mM EDTA/90 mM boric acid]. The prerun was for 30 min and the run for 3 h at 150 V and 4 °C in 1 × TBE. After electrophoresis the gel was dried and autoradiographed for 12–24 h. The supershift assay was performed in 10 μl of 20 mM Hepes, pH 7.9, containing 0.4 mM EDTA, 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 10% glycerol, 0.25 mM Pefabloc SC and 0.05% Nonidet P40. Human androgen receptor extract (1 μl of a 19.5 μg of protein/μl preparation) was incubated with 5α-dihydrotestosterone (1 μM) for 15 min on ice. Then 1 μl of ARP4 antiserum was added, which was followed by 1 h incubation on ice also. A mixture of 1 μl of the labelled annealed oligonucleotides (100000 c.p.m.) and 1 μl of 1 μg/1 μl poly[d(I-C)] was then added. After 1 h incubation at room temperature, the samples were analysed on a 4% polyacrylamide/bisacrylamide (29:1)/0.05% Nonidet P40/0.25 × TBE gel. The prerun was for 30 min and the run for 2.5 h at 140 V and 4 °C in 0.25 × TBE/0.05% Nonidet P40. Autoradiography of the dried gel was for 16 h.

RESULTS

Isolation of phage clones covering the *CRISP-1* gene

Since the 5′ regions of the *CRISP-1* and *CRISP-3* cDNAs are very conserved [7,8], the same *CRISP-3*-derived probe was used for the isolation of both genes. The positive phage clones obtained were further analysed with oligonucleotides able to discriminate between *CRISP-1* and *CRISP-3*, and also to identify those that harboured the longest gene fragment. Two of them, λ1 and λ16a, were thus selected for further characterization and subcloned into a plasmid vector after complete *Xho*II digestion. Bacterial clones carrying *CRISP-1* gene fragments were identified by colony hybridization. Analysis with oligonucleotide probes and by DNA sequencing revealed that the original λ1 phage harboured about 3.7 kb of *CRISP-1* promoter region and four exons, whereas the insert in λ16a had an additional fifth exon but was shorter upstream (Figure 1). Since the exon/intron organization was found to be similar to that of *CRISP-3* in this region [27], we used additional *CRISP-1*-specific probes likely to correspond to separate exons in the remainder of the gene to rescreen the genomic library. Three phage clones carrying exons 5 and 6 (λ43), 6 and 7 (λ16b) and 7 and 8 (λ9) of the *CRISP-1* gene were isolated in this way (Figure 1). Altogether the inserts of the five clones obtained encompassed the entire transcription unit of the *CRISP-1* gene.

Sequence and organization of the *CRISP-1* gene

The complete DNA sequence of the exons and flanking regions was determined. It showed that the *CRISP-1* gene was organized into eight exons and seven introns (Figure 1). The sequences of the exons were in complete agreement with that of the cDNA, except for the deletion of a T in the 3′-untranslated region (position 1268 in [8]). This is probably best explained by the different mouse strains from which the libraries were derived. The size of the internal exons varied between 74 and 146 bp (Figure 2) which puts them in the main category described [36]. The first exon was the shortest (see below) and coded for most of the 5′ untranslated part. The second exon coded for the entire 20 amino acid-long signal peptide and the four N-terminal amino acids of mature CRISP-1 [1]. The codons for the six cysteines linked pairwise to create a discrete N-terminal domain in CRISP-1 [1] were found in exons 4, 5 and 6. Exons 7 and 8 coded for the cysteine-rich C-terminal region of the protein. Exon 8 was much longer than the others but similar in size in its protein-coding

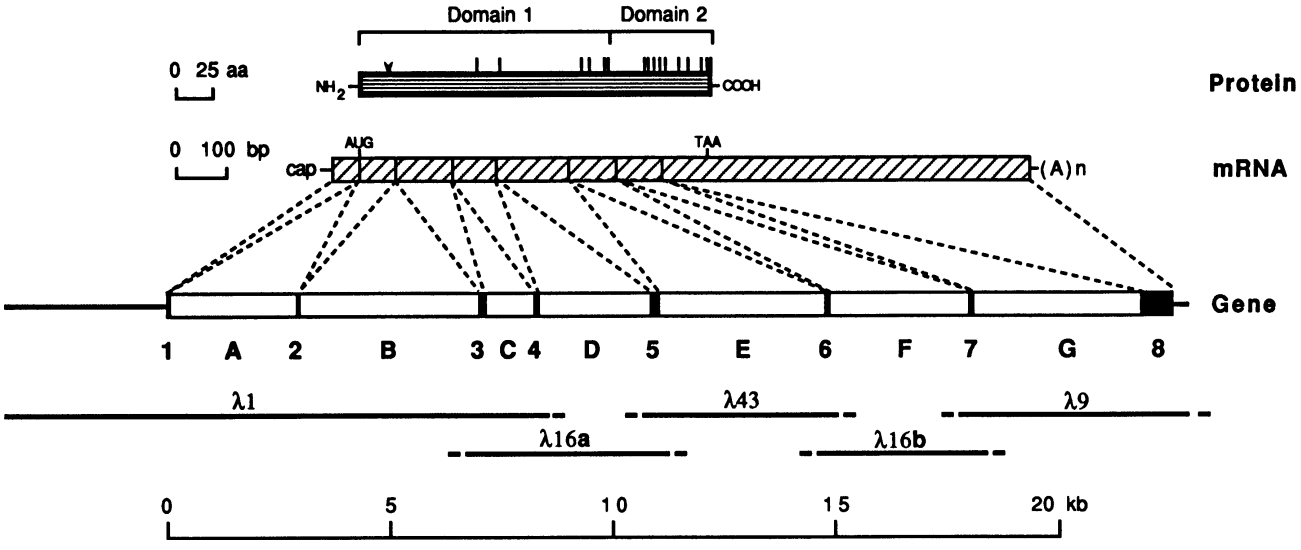

Figure 1 Organization of the *CRISP-1* gene

In the *CRISP-1* gene which is schematically drawn as an open bar, the exons are shown in black and marked 1–8, and the introns marked A–G. The five overlapping phage clones from which the sequence information was derived are indicated under the gene structure, with the corresponding scale. The top of the figure represents the protein (bar with horizontal lines) and its corresponding scale. The vertical arrowhead shows the limit between the signal sequence and the mature protein and small vertical lines indicate cysteine residues. The boundaries of the two domains are shown. *CRISP-1* mRNA represented immediately below (bar with diagonal lines), with its corresponding scale. The positions of the AUG and TAA translation-initiation and stop codons are indicated.

intron	EXON		intron	Phase
ATGTC..	**1**	(55 bp)	..CAGCAgtaagtag (~2.8 kb)	/
aaaaatcgtaatgcttgtctttgttctctttttgttctctagCCATG.. M1	**2**	(74 bp)	..CTCAGgtatgggt (~4.0 kb) Q24	0
tcatacactaatgtgactttcattttatttccctttagGAAAA.. E25	**3**	(117 bp)	..AAATGgtgagggc (~1.05 kb) M63	0
attattatttatagctatctcctttcctttttctctctagGAATG.. E64	**4**	(88 bp)	..AACTAgtatgtga (~2.45 kb) N93	1
tgctgccttaatgattggtctatttctttcccatagATTTA..	**5**	(146 bp)	..CTCAGgtaagaaa (~3.65 kb) Q141	0
ttttttctctgacacatgctccatattcattcctggtgtgcagGTTGT.. V142	**6**	(95 bp)	..CCTGTgtaagtat (~3.05 kb) V173	2
tcatctgatgattttgcatcttgaatttcagTGGCA..	**7**	(89 bp)	..ATGCAgtaagttt (~3.7 kb) T203	1
gaaaatgataatgaagtaatgaatatttttttctttatttcagCCAAT..	**8**	(747 bp)	..TTGGC	/

Figure 2 Exon–intron junctions of the *CRISP-1* gene

Exon and intron sequences are given in uppercase and lowercase letters respectively. The phase refers to the boundary relative to the last codon of an exon. The amino acids affected by the intron interruption are indicated by their respective location in the protein [8]. The branchpoint consensus motif near the 3′ end of each intron is underlined.

part to exons 2–7, in line with previous observations [36]. It also covered the entire 3′ untranslated part, including the polyadenylation signal.

Seven introns interrupted the *CRISP-1* gene (Figure 1). Introns B to G split the protein-coding regions into the three possible phases [37] with a majority of phase 0 introns (Figure 2). The exon/intron boundaries all conformed to the GT/AG splice rule [38] (Figure 2). A heptamer motif closely resembling the consensus found to be a branchpoint for the lariat formed during splicing [39–41] was found in the appropriate position upstream of the 3′ extremity of each intron (Figure 2). The intron size was determined by PCR (not shown) and found to vary between 1.05 and 4 kb. The entire *CRISP-1* gene therefore spanned over 20 kb of the mouse genome.

Determination of the transcription-initiation site

The transcription-initiation site was mapped by primer extension using a 30 nucleotide-long antisense oligonucleotide hybridizing near the 5′ end of the *CRISP-1* cDNA and epididymis poly(A)$^+$ RNA as template. Two major bands differing by one nucleotide in length and a shorter much fainter band were generated (Figure 3). The longest and most abundant elongation product ended 57 bp upstream of the translation-initiation codon and 32 bp upstream of the 5′ extremity of the published *CRISP-1* cDNA [8]. This position is occupied by an adenine residue embedded in a region, in good agreement with the consensus cap site sequence [42] and at a correct distance downstream of the TATA box (see below). It was taken as nucleotide 1 in the genomic sequence (Figure 4). Altogether these results are compatible with the previously determined size of the *CRISP-1* mRNA and indicate that the first exon is made of 55 bp of 5′-untranslated sequence. The transcription-start site of *CRISP-3* was mapped at an equivalent position [27].

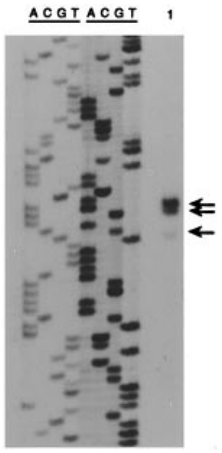

Figure 3 Primer extension mapping of the transcription-initiation point

Mouse epididymis poly(A)$^+$ RNA (2 μg; lane 1) was used as a template for reverse transcription. The arrows point to the extension products. Two different sequencing reactions (lanes ACGT) served as markers for a precise determination of the length. Exposure was for 48 h.

Characterization of the promoter region

A 3714 bp-long stretch of 5′ upstream region was sequenced (Figure 4). A TATA box was detected at the appropriate position with regard to the transcription-initiation site (positions −29 to −22) whereas no sequence exactly matching the CCAAT box was identified in the vicinity [43]. Several stretches resembling the ARE consensus sequence GGWACANNNTGTTCT in either orientation [44] were found (Figure 4, boxes). Two of them with one and two mismatches to the consensus were spaced by only

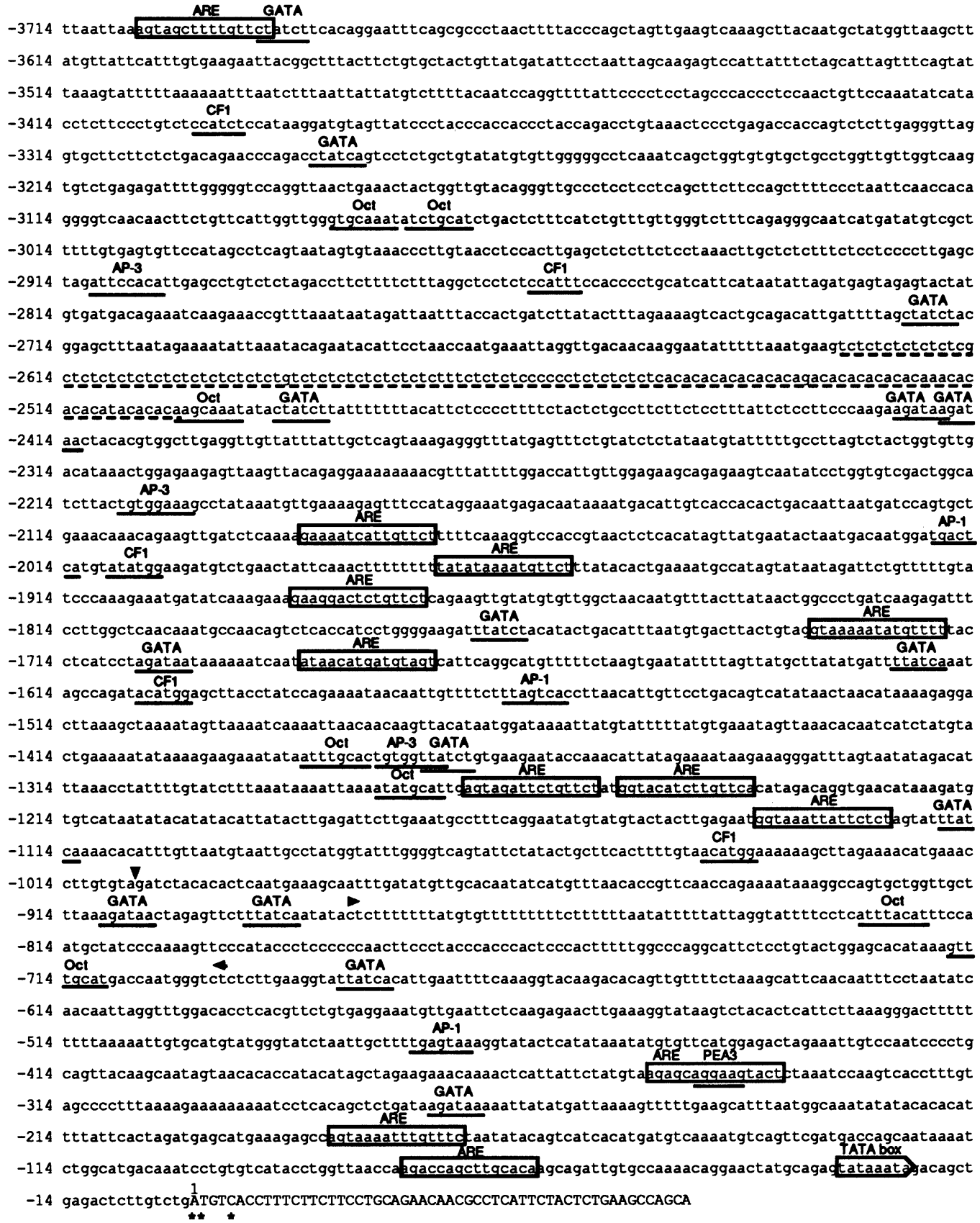

Figure 4 Nucleotide sequence of the *CRISP-1* promoter region

The DNA sequence of the 5′ upstream region (lowercase letters) and of the first exon (uppercase letters) is shown. Asterisks indicate the transcription-initiation sites as determined by primer extension. The TATA box is shown in an arrow. Sequences matching the following motifs in either orientation are highlighted in boxes (AREs) or by underlining: GGWACANNNTGTTCT (ARE); ATGCAAAT (Oct); MGGAAGT (PEA3); WGATAR (GATA); TGASTMA (AP-1); TGTGGWWW (AP-3); ANATGG (CF1). Dotted underlining indicates poly[d(C-A)] and [d(C-T)] stretches. A vertical arrowhead shows the position corresponding to the 5′ end of the published *CRISP-3* upstream region and converging arrowheads delineate the insertion found in the *CRISP-1* promoter. The GeneBank/EMBL accession number for this sequence is Y09162.

Table 1 Putative AREs found in the *CRISP-1* and *CRISP-3* promoters

The sequences were oriented to obtain the best fit with the TGTTCT ARE half-site. Motifs with three or fewer mismatches to the entire consensus are shown. For *CRISP-1*, the numbering refers to Figure 4 and gives the position of the first nucleotide with regard to the transcription-start site. The *CRISP-3* elements starting at −369 and −171 are taken from [27] whereas the −720 and −335 elements were found by new analysis of the sequence. An asterisk indicates that the complementary strand is shown. Underlined positions diverge from the consensus [44] given at the bottom.

Gene	Position	Sequence
CRISP-1	−3706	AGTAGCtttTGTTCT
	−2088	GAAAATcatTGTTCT
	−1973	TATATAaaaTGTTCT
	−1889	GAAGGActcTGTTCT
	−1732	GTAAAataTGTTTT
	−1674*	ACTACAtcaTGTTAT
	−1270	AGTAGAttcTGTTCT
	−1253	GGTACAtctTGTTCA
	−1138	GGTAAAttaTTCTCT
	−336*	AGTACTtccTGCTCT
	−171*	GAAACAaatTTTACT
	−63*	TGTGCAagcTGGTCT
CRISP-3	−720	GGTAACtggAGTTCT
	−369	AAAACTcatTGTTCT
	−335*	AGTACTtcaTGCTCT
	−171*	GAAACAaatTTTACT
ARE consensus		GGAACAnnnTGTTCT T

2 bp (starting at positions −1270 and −1253). Ten other putative AREs with three mismatches to the consensus were also found. None of these sequences formed a perfect dyad symmetry as observed for other AREs [44]. A comparison of the putative AREs of the *CRISP-1* and *CRISP-3* promoter regions is given in Table 1.

Several other motifs possibly recognized by transcription factors [45] were found (Figure 4). Seven regions matching the consensus bound by the octamer transcription factors (Oct or OTF) in seven of eight positions were detected as well as multiple GATA factor-recognition sites and five CF1 sites. In addition AP-1 and AP-3 motifs were present three times each and a consensus PEA3 sequence was detected in the ARE starting at −350.

Another remarkable feature of the *CRISP-1* promoter is the existence of a poly[d(C-A)] and poly[d(C-T)] stretch (Figure 4; between positions −2629 and −2502). The $[d(C\text{-}A)]_n$ sequence might play a role in gene regulation because of its potential to adopt the Z-DNA conformation and facilitate the binding of regulatory proteins to neighbouring *cis* elements [46,47].

Characterization of androgen receptor–ARE complexes

Human androgen receptor-containing extracts were prepared from infected Sf9 cells and analysed by Western blot. The ARP4 antiserum raised against a peptide derived from the N-terminal transactivation domain of the human androgen receptor [35] was used. A major band of 110 kDa was observed for the cells infected with recombinant baculovirus, confirming the presence of androgen receptor in these extracts (Figure 5). Several smaller much fainter bands possibly corresponding to degradation products were also seen and a similar pattern has also been observed by others [32]. No signal was visible for control extracts from Sf9 cells infected with wild-type viral DNA. For EMSA, the complementary oligonucleotides BS1 and BS2, which cover the

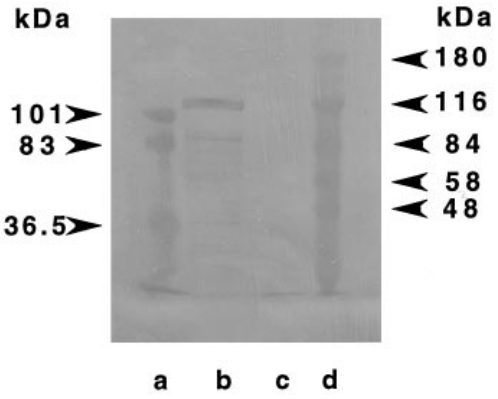

Figure 5 Western-blot analysis of infected Sf9 cell extracts

Sf9 cell extracts were electrophoresed on a 7.5% polyacrylamide gel and transferred on to a BA85 nitrocellulose membrane which was analysed using a 1:1000 dilution of the ARP4 antiserum. Lanes a and d, molecular-mass markers; lane b, extract from Sf9 cells infected with recombinant baculovirus expressing the androgen receptor; lane c, extract from Sf9 cells infected with wild-type viral DNA.

CRISP-1 AREs starting at −1270 and −1253, were synthesized. These AREs are the most closely related to the consensus in the *CRISP-1* promoter and they are spaced by 2 bp only. The BS1/BS2 probe was incubated with androgen-receptor-containing extracts, and a strong signal was obtained. Addition of unlabelled competitor oligonucleotides clearly reduced the intensity of this signal, indicating that the interaction was specific (Figure 6a, lanes 1–4). No complex was formed when a control extract without androgen receptor was used (lane 13). Shorter oligonucleotides were than devised reproducing either ARE while modifying one of the flanking regions so as not to include the partial ARE sequence from the closely neighbouring element. BS5 and BS6 corresponded to the −1253 element and BS9 and BS10 to the −1270 element. We found that BS5/BS6 and BS9/BS10 were both able to interfere with the complex-formation between the BS1/BS2 probe and the androgen receptor, but BS9/BS10 competed less well, as evidenced when using a 500-fold excess (lanes 5–10). An unrelated competitor oligonucleotide pair interfered far less with complex-formation (lanes 11 and 12). We next compared the ability of the −1270 and −1253 AREs to individually bind to the androgen receptor, using labelled BS5/BS6 and BS9/BS10 as probes (Figure 6b). Formation of complexes with androgen-receptor-containing extracts was observed in all cases, but the amount of retarded DNA was far less in the case of BS9/BS10 compared with BS5/BS6 or BS1/BS2. The extent of mobility shift was similar for the three probes used. In order to confirm the authenticity of the complex formed further, a supershift experiment was carried out using the ARP4 antiserum [35]. The androgen receptor extract was incubated with the antiserum before adding the BS5/BS6 probe. The complex formed was analysed using a different gel concentration from before for better visualization, and indeed a supershift was observed for the upper band which had previously been identified as the specific complex (Figure 6c).

DISCUSSION

The structure of the murine *CRISP-1* gene is very similar to that of the previously characterized *CRISP-2/Tpx-1* and *CRISP-3*

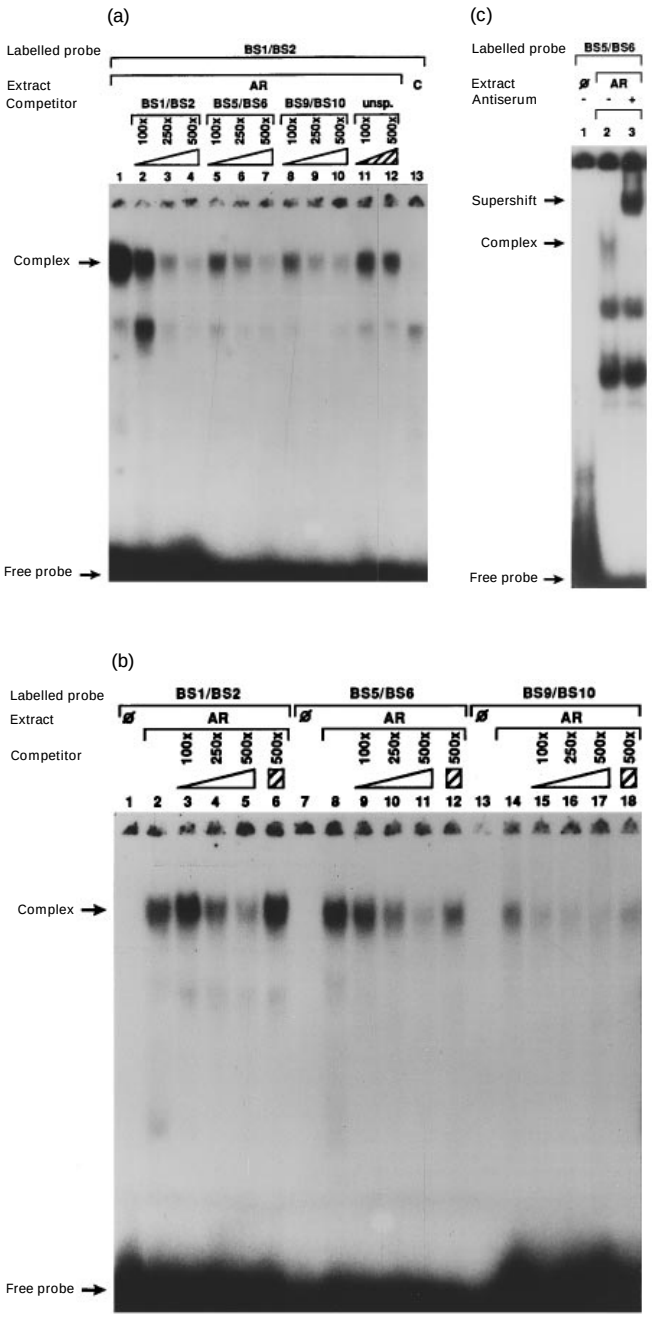

Figure 6 Binding of *CRISP-1* AREs to the androgen receptor

EMSA was performed after incubation of androgen-receptor-containing extracts prepared from infected Sf9 cells with double-stranded labelled probes. The arrows point to the androgen-receptor-specific complex and to the free probe. (**a**) Complex-formation between the BS1/BS2 probe which covers the −1270 and −1253 AREs and an androgen receptor extract (AR, lanes 1–12) or a control extract (C, lane 13) was analysed. The indicated molar excesses of BS1/BS2 (lanes 2–4), BS5/BS6 (corresponding to the −1253 ARE; lanes 5–7), BS9/BS10 (corresponding to the −1270 ARE; lanes 8–10) and of an unspecific oligonucleotide pair (unsp., lanes 11 and 12) were added for competition. (**b**) BS1/BS2 (lanes 1–6), BS5/BS6 (lanes 7–12) and BS9/BS10 (lanes 13–18) were used as probes. Lanes 1, 7, 13, free probes (∅); lanes 2, 8, 14, probes in the presence of androgen receptor extract (AR); lanes 3–5, 9–11, 15–17, probes in the presence of androgen receptor extract and the indicated molar excesses of unlabelled specific competitor; lanes 6, 12, 18, probes in the presence of androgen receptor extract and a 500-fold excess of unlabelled unspecific competitor. (**c**) BS5/BS6 was used as probe. Lane 1, free probe; lane 2, probe in the presence of androgen receptor extract; lane 3, probe in the presence of androgen receptor extract and the ARP4 anti-(androgen receptor) serum.

genes. The protein-coding parts are interrupted at equivalent positions and in the same phases by introns. Several features indicate that the divergence between *CRISP-1* and *CRISP-3* is more recent than the divergence between *CRISP-2/Tpx-1* and the *CRISP-1/CRISP-3* ancestor. The promoter regions of *CRISP-1* and *CRISP-3* are almost colinear in the region spanning from −696 to the transcription-start site (only two insertions in the *CRISP-1* promoter) where they share 92% identity. The remainder of the published *CRISP-3* promoter is 84.5% identical with the −1006 to −884 region of *CRISP-1*, suggesting that an insertion or deletion event accounts for the −883 to −697 region of the *CRISP-1* promoter (Figure 4). No significant sequence identity was found with the *CRISP-2/Tpx-1* promoter. Additional remarkable differences between *CRISP-2/Tpx-1* on the one hand and *CRISP-1* and *CRISP-3* on the other are the presence of two additional untranslated exons in the 5′ region of *CRISP-2/Tpx-1*, and of an unusually large penultimate intron (approximately 17 kb) which separates the region coding for the cysteine-rich C-terminal domain from the rest of the gene. This closer evolutionary relationship between *CRISP-1* and *CRISP-3* is also reflected by their parallel, albeit differential, androgen regulation [7,8] and their higher coding DNA sequence identity.

Several stretches with close similarity to the ARE consensus were detected in the *CRISP-1* promoter region. The two that differed least from the consensus were analysed and found to bind to the androgen receptor *in vitro* as a tandem, with the more proximal element playing the main role in the interaction observed. Since the complex formed by the probe covering the tandem AREs migrated with essentially the same mobility as the one formed by the individual AREs, it is unlikely that the two elements are bound concomitantly. This is probably explained by both the weak binding affinity exhibited by the distal element and the close spacing between these AREs. A similar situation has been reported for the *Slp* gene promoter which contains three closely spaced AREs of which one is essential for hormone responsiveness [48,49]. Additional studies are needed to investigate the other less consensual ARE motifs found in the *CRISP-1* promoter. Preliminary results indicate that several of them are indeed able to bind weakly to the androgen receptor (not shown).

Seven Oct (or OTF) motifs, all with one deviation from the consensus, were observed in the *CRISP-1* promoter. They are all further upstream than the two Oct motifs of the *CRISP-3* promoter that have recently been shown to be targets for the B-cell transcription factor Oct-2 [22]. Remarkably, these two motifs are mutated at one or two positions in *CRISP-1* and are therefore probably no longer recognized as such. This probably explains why *CRISP-3* (and not *CRISP-1*) is expressed under Oct-2 control in pre-B cells, and suggests that the upstream Oct elements of *CRISP-1* are bound by the ubiquitous transcription factor Oct-1, if they are bound at all. Co-operative interactions of Oct-1 with the progesterone and glucocorticoid receptors in the context of the induction of the mouse mammary tumour virus promoter and with the androgen receptor for the induction of the *Slp* gene have been described [50,51].

The presence of a PEA3 motif is particularly interesting because of the tissue-restricted expression of the cognate transcription factor, which has only been detected in the epididymis and brain [52]. Several such motifs have also been found in two other genes with strong epididymal expression, namely those coding for the androgen-dependent glutathionine peroxidase-like protein arMEP24 [53] and the β-hexosaminidase [54]. Other putative *cis* elements found in *CRISP-1* are recognized by more general transcription factors such as AP-1, AP-3 and CF1, and they might act in co-operation with the androgen receptor, as has

been shown in the case of the mouse mammary tumour virus promoter [55].

The presence of numerous AREs with different binding affinities for the androgen receptor is probably important for the regulation of the *CRISP-1* gene. Binding sites that are able to sense different hormone concentrations might be critical for the fine tuning of gene expression as has already been pointed out [48]. Their occurrence suggests complex interactions in which the transcription factors likely to interact with the other *cis* elements found in the promoter are probably also involved. Deciphering the complex interactions underlying the androgen control as well as identifying the regions and coactivators involved in the tissue-restricted expression of the *CRISP* genes will require additional studies for which the *CRISP-1* gene presented here should represent a valuable model.

We thank Professor Andrew Cato for supplying the human androgen receptor cDNA clone and Dr. Thomas Petri and Brigitte Ockert for setting up the androgen receptor expression system in insect cells. We are deeply indebted to Professor Olli Jänne and Dr. Jorma Palvimo for the gift of androgen receptor antiserum. The help of Dr. Johanna Beekman with the EMSA is gratefully acknowledged. We also thank Dr. Jörn Krätzschmar for careful reading of the manuscript. We are indebted to Iris Schüttke for excellent technical assistance and to Daniela Sauvageot for the supply of numerous oligodeoxynucleotides. This work was supported in part by grant 0310681 from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (Jülich, Germany).

REFERENCES

1 Eberspaecher, U., Roosterman, D., Krätzschmar, J., Haendler, B., Habenicht, U.-F., Becker, A., Quensel, C., Petri, T., Schleuning, W.-D. and Donner, P. (1995) Mol. Reprod. Dev. **42**, 157–172
2 Cooper, T. G. (1990) Fertil. Steril. **54**, 965–975
3 Robaire, B. and Viger, R. S. (1995) Biol. Reprod. **52**, 226–236
4 Ariel, M., Cedar, H. and McCarrey, J. (1994) Nature Genet. **7**, 59–63
5 Rankin, T. L., Tsuruta, K. J., Holland, M. K., Griswold, M. D. and Orgebin-Crist, M.-C. (1992) Biol. Reprod. **46**, 747–766
6 Vierula, M. E., Rankin, T. L. and Orgebin-Crist, M.-C. (1995) Microsc. Res. Technique **30**, 24–36
7 Mizuki, N. and Kasahara, M. (1992) Mol. Cell. Endocrinol. **89**, 25–32
8 Haendler, B., Krätzschmar, J., Theuring, F. and Schleuning, W.-D. (1993) Endocrinology **133**, 192–198
9 Krätzschmar, J., Haendler, B., Eberspaecher, U., Roosterman, D., Donner, P. and Schleuning, W.-D. (1996) Eur. J. Biochem. **236**, 827–836
10 Hayashi, M., Fujimoto, S., Takano, H., Ushiki, T., Abe, K., Ishikura, H., Yoshida, M. C., Kirchhoff, C., Ishibashi, T. and Kasahara, M. (1996) Genomics **32**, 367–374
11 Cameo, M. S. and Blaquier, J. A. (1976) J. Endocrinol. **69**, 47–55
12 Brooks, D. E. (1987) Ann. N. Y. Acad. Sci. **513**, 179–194
13 Brooks, D. E., Means, A. R., Wright, E. J., Singh, S. P. and Tiver, K. K. (1986) Eur. J. Biochem. **161**, 13–18
14 Brooks, D. E. and Tiver, K. (1983) J. Reprod. Fertil. **69**, 651–657
15 Vreeburg, J. T. M., Holland, M. K. and Orgebin-Crist, M.-C. (1992) Biol. Reprod. **47**, 588–597
16 Fournier-Delpech, S., Courot, M. and Dubois, M. P. (1985) J. Androl. **6**, 246–250
17 Rochwerger, L., Cohen, D. J. and Cuasnicú, P. S. (1992) Dev. Biol. **153**, 83–90
18 Perez Martinez, S., Conesa, D. and Cuasnicú, P. S. (1995) J. Reprod. Immunol. **29**, 31–45
19 Kasahara, M., Gutknecht, J., Brew, K., Spurr, N. and Goodfellow, P. N. (1989) Genomics **5**, 527–534
20 Hardy, D. M., Huang, T. T. F., Driscoll, W. J., Tung, K. S. K. and Wild, G. C. (1988) Biol. Reprod. **38**, 423–437
21 Foster, J. A. and Gerton, G. L. (1996) Mol. Reprod. Dev. **44**, 221–229
22 Pfisterer, P., König, H., Hess, J., Lipowsky, G., Haendler, B., Schleuning, W.-D. and Wirth, T. (1996) Mol. Cell. Biol. **16**, 6160–6168
23 Kjeldsen, L., Cowland, J. B., Johnsen, A. H. and Borregaard, N. (1996) FEBS Lett. **380**, 246–250
24 Rigden, J. and Coutts, R. (1988) Trends Genet **4**, 87–89
25 Mochca-Morales, J., Martin, B. M. and Possani, L. D. (1990) Toxicon **28**, 299–309
26 Morrissette, J., Krätzschmar, J., Haendler, B., El-Hayek, R., Mochca-Morales, J., Martin, B. M., Patel, J. R., Moss, R. L., Schleuning, W.-D., Coronado, R. and Possani, L. D. (1995) Biophys. J. **68**, 2280–2288
27 Schwidetzky, U., Haendler, B. and Schleuning, W.-D. (1995) Biochem. J. **309**, 831–836
28 Kasahara, M., Hayashi, M., Yoshida, M. C., Nadeau, J. H., Fujimoto, S. and Ishibashi, T. (1995) Mamm. Genome **6**, 52–54
29 Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395
30 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
31 Culig, Z., Hobisch, A., Cronauer, M. V., Cato, A. C. B., Hittmaier, A., Radmayr, C., Eberle, J., Bartsch, G. and Klocker, H. (1993) Mol. Endocrinol. **7**, 1541–1550
32 Xie, Y.-B., Sui, Y.-P., Shan, L.-X., Palvimo, J. J., Phillips, D. M. and Jänne, O. A. (1992) J. Biol. Chem. **267**, 4939–4948
33 Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
34 Blake, M. S., Johnston, K. H., Russell-Jones, G. J. and Gotschlich, E. C. (1984) Anal. Biochem. **136**, 175–179
35 Jänne, O. A., Palvimo, J. J., Kallio, P. and Mehto, M. (1993) Ann. Med. **25**, 83–89
36 Traut, T. W. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2944–2948
37 Patthy, L. (1987) FEBS Lett. **214**, 1–7
38 Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kédinger, C. and Chambon, P. (1980) Science **209**, 1406–1414
39 Keller, E. B. and Noon, W. A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7417–7420
40 Zeitlin, S. and Efstratiadis, A. (1984) Cell **39**, 589–602
41 Montgomery, J. C., Venta, P. J., Eddy, R. L., Fukushima, Y.-S., Shows, T. B. and Tashian, R. E. (1991) Genomics **11**, 835–848
42 Xu, L., Thali, M. and Schaffner, W. (1991) Nucleic Acids Res. **19**, 6699–6704
43 Bucher, P. and Trifonov, E. N. (1986) Nucleic Acids Res. **14**, 10009–10026
44 Roche, P. J., Hoare, S. A. and Parker, M. G. (1992) Mol. Endocrinol. **6**, 2229–2235
45 Faisst, S. and Meyer, S. (1992) Nucleic Acids Res **29**, 3–26
46 Rich, A., Nordheim, A. and Wang, A. H.-J. (1984) Annu. Rev. Biochem. **53**, 791–846
47 Lancillotti, F., Lopez, M. C., Arias, P. and Alonso, C. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 1560–1564
48 Adler, A. J., Scheller, A., Hoffman, Y. and Robins, D. M. (1991) Mol. Endocrinol. **5**, 1587–1596
49 Adler, A. J., Scheller, A. and Robins, D. M. (1993) Mol. Cell. Biol. **13**, 6326–6335
50 Brüggemeier, U., Kalff, M., Franke, S., Scheidereit, C. and Beato, M. (1991) Cell **64**, 565–572
51 Scarlett, C. O. and Robins, D. M. (1995) Mol. Endocrinol. **9**, 413–423
52 Xin, J.-H., Cowie, A., Lachance, P. and Hassell, J. A. (1992) Genes Dev. **6**, 481–496
53 Ghyselinck, N. B., Dufaure, I., Lareyre, J.-J., Rigaudière, N., Mattéi, M.-G. and Dufaure, J.-P. (1993) Mol. Endocrinol. **7**, 258–272
54 Triggs-Raine, B. L., Benoit, G., Salo, T. J., Trasler, J. M. and Gravel, R. A. (1994) Biochim. Biophys. Acta **1227**, 79–86
55 Cato, A. C. B., Skroch, P., Weinmann, J., Butkeraitis, P. and Ponta, H. (1988) EMBO J. **7**, 1403–1410

Received 7 May 1996/12 August 1996; accepted 6 September 1996