Characterization of a Mammalian Gene Related to the Yeast CCR4 General Transcription Factor and Revealed by Transposon Insertion*

Anne Dupressoir, Willy Barbot, Marie-Paule Loireau, and Thierry Heidmann‡

From the Unité des Rétrovirus Endogènes et Éléments Rétroïdes des Eucaryotes Supérieurs, CNRS UMR 1573, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France

The general transcription factor CCR4 from Saccharomyces cerevisiae (yCCR4) is required for the transcription of several nonfermentative genes, including that of the glucose-repressible alcohol dehydrogenase (1, 2). It is part of a multi-subunit complex, several components of which have been characterized including DBF2, a cell cycle-regulated protein kinase (3), and proteins of the NOT family, involved in repression of transcription (4). yCCR4 displays a glutamine-rich N terminus, like many factors involved in transcriptional activation, and a leucine-rich repeat motif that is necessary for protein-protein interactions within the complex (5, 6). Both motifs are required for transcriptional activation by yCCR4. The yCCR4 complex was demonstrated to function at a post-chromatin remodeling step in the case of the ADH2 gene (7) and to play an important role in several cellular processes: yCCR4 affects the expression of genes involved in cell wall integrity (3), in UV sensitivity (8), and in methionine biosynthesis (9). Although the mechanism of how yCCR4 functions remains undefined (yet the yCCR4 protein does not bind DNA by itself, Ref. 6), several lines of evidence suggest that the complex plays an important and conserved role in vertebrate gene regulation as well. Actually, there is increasing evidence for the existence of yCCR4-like proteins in higher eucaryotes. First, a gene encoding a yCCR4-like protein (named nocturnin) has been identified in Xenopus (10). Nocturnin displays a strong similarity to the C-terminal part of yCCR4 and contains a leucine zipper-like dimerization domain. Interestingly, nocturnin expression discloses a strong circadian regulation in the retina, suggesting that it could be involved in the circadian clock system. Second, the existence of yCCR4-like proteins in mammals is strongly supported by the results of Draper et al. (11). Those authors identified a mouse protein (mCAF1) which can functionally interact, in a yeast two-hybrid assay, with the yeast yCCR4 factor, thus indicating a strong evolutionary conservation of at least some components of a CCR4-containing transcriptional regulatory complex. Finally, our group identified, at the 3′-end of an intracisternal A-particle chimeric murine transcript, a 1-kb\(^1\) open reading frame (ORF) with significant similarity to the yCCR4 C-terminal domain (12). Intracisternal A-particle (IAP) sequences are moderately reiterated transposable elements (~1000 copies in the mouse genome) which are closely related to retroviruses and transpose via the reverse transcription of an RNA intermediate (13, 14). They are flanked by two long terminal repeats (LTR), with a U3-R-U5 organization, that contain the signals for the initiation and regulation of transcription (5′-LTR) and for the polyadenylation of the transcripts (3′-LTR). IAPs are one of the most potent insertional mutagens in the mouse, and it is well documented that their insertion can perturb the normal pattern of expression of the targeted cellular genes (reviewed in Refs. 13 and 15). In this respect, we had previously characterized three IAP transcripts of abnormal size (3, 6, and 10 kb, named IAP-AR transcripts), which are induced exclusively in the liver of old mice (16). We have shown that the 10-kb IAP-AR transcript is initiated within the 5′-LTR of an...
IAP gene and corresponds to a transcriptional read-through extending beyond the 3′-LTR into a cellular sequence containing the 1-kb ORF with significant similarity to the γCCR4 C-terminal domain (12). This 10-kb transcript gives rise to the two smaller transcripts by one or two splicing events using donor and acceptor sites within the IAP sequence and an acceptor site 5′ to the γCCR4-related ORF. We hypothesized that the IAP sequence was inserted into an intron of a mouse gene related to the γCCR4 gene. As a first step to elucidate the potential role of this factor in mammals, we have now characterized this gene, tentatively named mCCR4. We present an analysis of its genomic organization and of the structure of its associated transcripts. mCCR4 has all the characteristics features of a “housekeeping gene,” with a CpG island-containing promoter and a ubiquitous expression in the mouse. We further show that the IAP insertion, in the first intron, is a recent phenomenon, being absent in some laboratory mouse strains. A comparison of the mCCR4 transcripts in both types of mice discloses no effect of the IAP insertion on the structure and pattern of expression of the mCCR4 gene (with the exception of the liver of aged mice, see above). This rather unusual situation suggests a strong selective pressure for maintaining mCCR4 function, a result consistent with the high amino acid conservation found with the human homologue of mCCR4, which we have also characterized in the present work, and a central role for these factors in vertebrates.

EXPERIMENTAL PROCEDURES

Characterization of Murine and Human γCCR4-related cDNAs—Cloning of the murine γCCR4-related cDNA was achieved by 5′- and 3′-rapid amplification of cDNA ends (RACE) using a Marathon-Ready™ mouse brain cDNA library (CLONTECH, constructed from normal whole brains pooled from 9–11-week old BALB/c males). Priming complementary to the adapter sequence and various primers derived from the mCCR4 sequence were used in nested PCR reactions. For exons 2 and 3 of the human γCCR4-related cDNA, we used a Marathon-Ready™ human placenta cDNA library (CLONTECH) and low-stringency PCR reactions (a first step at 94 °C for 60 s; then 4 cycles at 94 °C for 30 s and 61 °C for 50 s; 4 cycles at 94 °C for 20 s and 59 °C for 50 s; 4 cycles at 94 °C for 20 s and 57 °C for 50 s; 30 cycles at 94 °C for 20 s, 55 °C for 30 s, and 72 °C for 60 s; and ending with a step at 72 °C for 2 min), Characterization of the 5′-end of the human cDNA (exon 1), which was achieved by oligo(dT)-primed human cDNA library2 in which the cDNAs, prepared from mRNA isolated from NTera2D1 teratocarcinoma cells, are inserted into a retroviral vector. Successive 5′-RACE PCR reactions were performed using primers in the human sequence and primers in the retroviral vector. All PCR reactions, except the low-stringency PCRs, were performed using the Advantage-GC CDNA Polymerase Mix (CLONTECH, specifically designed for GC-rich cDNAs) with a 95 °C denaturation step for 80 s followed by 60 cycles at 95 °C for 20 s, 66 °C for 3 min, and final extension at 66 °C for 10 min. Amplification products were subcloned into the T-vector (Promega) and sequenced.

Characterization of Genomic DNA—PCR amplifications with genomic DNA were performed using the Advantage-GC CDNA Polymerase Mix (CLONTECH). Mapping of exon-intron boundaries was achieved by PCR amplification of intrinsic sequences with primers designed from the cdNA sequences. For large introns, long distance PCR was performed after a 95 °C denaturation step for 2 min, by 60 cycles at 95 °C for 20 s, and 66 °C for 14 min. Final extension was at 66 °C for 10 min. Inverted PCR amplification was performed as described in Ref. 14 using genomic DNA first digested by restriction enzymes and then self-ligated, and pairs of primers in opposite orientation. A mouse 129/sv lambda genomic library (Stratagene) was screened using a PCR-based method (17) with primers bracketing the exon I/intron 1 junction. 105 plaque-forming units were amplified in an XL1-Blue MRA host strain, distributed in 96-well plates, and then subjected to PCR amplification. 200 plaque-forming units of a single positive well were further analyzed using a conventional plaque hybridization assay, leading to the isolation of an mCCR4-containing phage.

DNA Isolation and Southern Blot Analysis—High molecular weight DNA was extracted from the tail of various inbred strains. Tails were lysed overnight in 50 mM Tris (pH 8), 10 mM EDTA (pH 8), 100 mM NaCl, 0.2% SDS and 500 μg/ml Proteinase K. DNA was then extracted by phenol-chloroform treatment and ethanol precipitation. For each restriction analysis, 10 μg of DNA was cleaved with at least a 4-fold excess of restriction enzyme for 10 h, electrophoresed on a 1% agarose gel, and then analyzed by Southern blotting (18). A SpeI-BstXI fragment (460 bp) isolated from an inverted PCR product containing the region 5′ to the IAP was first used as a probe. After dehybridization, the same blot was then rehybridized with an intronic probe generated by PCR amplification of mouse genomic DNA using primers AR-B11 and AR-B12 (probe A in Ref. 12).

Assay for Promoter Activity—The pGL3 expression vector (Promega) contains the firefly luciferase as a reporter gene was used to assay the promoter activity of cloned genomic DNA fragments from the 5′-region of the mCCR4 gene. Several fragments were excised from an inverted PCR product containing the 5′-region of the mCCR4 gene, blunt-ended, and cloned bidirectionally into the Smal site of the pGL3 vector: a XbaI-BglII fragment, [-500/+114] pGL3 and [+114/+500] pGL3 constructs; a XmaI-BglII fragment, [-227/+114] pGL3 and [+114/+227] pGL3 constructs an AluNI-BglII fragment, [-100/+114] pGL3, and [+114/+100] pGL3 constructs; and a DraIII-BglII fragment (+1/4/+114] pGL3 and [+114/+4] pGL3 constructs. The [BssHII-I] pGL3 construct (+30/+114] pGL3) was subsequently generated by deleting the BssHII-XbaI fragment from the [BglII-I] pGL3 construct. A [BssHII-I] pGL3 construct (+30/+114] pGL3 was obtained by deleting the BssHII-HindIII fragment from the [BglII-XbaI] pGL3 vector. A basic pGL3 promoterless luciferase vector and an SV40 promoter/enhancer luciferase vector were used as negative and positive control vectors, respectively. NIH-3T3 cells were grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum and transfected using the LipofectAMINE PLUS™ Reagent (Life Technologies, Inc.). Cells (5 × 104) were transfected with 1 μg of luciferase construct and 1 ng of a β-galactosidase containing vector (CMVβ, Stratagene) as a control for transfection efficiency. Two days post-transfection, cells were assayed for luciferase using Reporter Lysis Buffer (Promega) and for β-galactosidase using CPRG (Roche Molecular Biochemicals).

RNA Isolation and Northern Blot Analyses—Total cellular RNAs from various tissues were extracted using a pre-packed spun column containing silica gel-based membrane (RNeasy kit, QIAGEN). For Northern blot analysis, 15 μg of total RNA/lane were fractionated on a 1% agarose/formaldehyde gels. RNAs were transferred to a nylon charged membrane, RNAs upon UV illumination of the membrane and hybridization of the blots with a β-actin probe. Prehybridization and hybridization were performed in 7% SDS, 1 mM EDTA, 0.5 μM Na2HPO4. A PCR fragment (1.2 kb) containing exons 2 and 3 from mCCR4 cDNA was used as a probe. Hybridized blots were washed twice in 0.5% SSC, 0.1% SDS at 65 °C for 30 min. Filters were then exposed for at least 3 days.

RESULTS

Murine and Human γCCR4-related cDNAs—A 966-bp open reading frame with similarity to the yeast CCR4 gene was previously identified at the 3′-end of IAP LTR-initiated chimeric transcripts (IAP-AR) resulting from a read-through into flanking cellular sequences and detected exclusively in the liver of old mice (12). These included a 10-kb transcript and two processed transcripts of 6 and 3 kb (Fig. 1A). We hypothesized this ORF to be part of the coding region of a putative murine homologue, mCCR4, of the yeast CCR4 gene. To characterize mCCR4-specific transcripts not initiated within the IAP LTR but initiated, possibly, within the mCCR4 promoter, we used RNA from the brain as (i) no chimeric, IAP-promoted, transcript was ever been detected in this organ (Ref. 16, and Fig. 1B), and (ii) this organ expresses a high level of mCCR4-related transcripts (Fig. 1B and extensive analysis in Fig. 6). A pre-made Marathon-Ready brain cDNA library (CLONTECH) was therefore used for 5′- and 3′-RACE, with primers designed based on the previously characterized ORF sequence (see Fig. 1B for the strategy of cloning). 3′-RACE PCR amplified a 1.5-kb fragment, which was found to contain the 3′-end of the previ-
A Northern blot of transcripts from the liver (L) of aged mice is shown on the left, using the indicated 966-bp CCR4 ORF probe for hybridization. The 10-, 6-, and 3-kb transcripts are indicated on the left and their structure schematized on the diagram. B, non-IAP CCR4-related gene transcripts from mouse brain (Br) and strategy for cloning the corresponding cDNAs. A Northern blot is shown on the left (same conditions as in panel A). The sequences in common with the IAP-CCR4 chimeric transcripts are schematized as in panel A, with the two initiation sites indicated by arrows. Gene-specific primers (filled triangles) used for RACE PCR, and PCR products obtained using each of these primers together with the Marathon library adaptor-specific primers (open triangles) are indicated.

We also looked for possible human homologues of these mCCR4 transcripts. Previous data base searches had revealed several human expressed sequence tags that exhibited high levels of similarity to the 3′-region of the murine CCR4 and the Xenopus nocturnin cDNAs (10, 12). We therefore performed low-stringency PCR amplifications using a human pre-made Marathon-Ready placenta cDNA library (CLONTECH), with one primer derived from the T87026 human expressed sequence tag sequence and a second primer derived from the murine CCR4 3′-coding regions to identify the 3′-part of the human homologue (exons 2 and 3). As previously observed for the murine sequences, cloning of the cDNA 5′-region was rendered difficult by its high G/C content, at a position similar to that of the murine gene. We performed several successive 5′-RACE PCRs using an oligo(dT)-primed human cDNA library cloned into a retroviral vector, with primers derived from human and vector sequences. We finally obtained a partial hCCR4 cDNA sequence of 2397 bp. This cDNA contains a complete ORF of 1317 nt that could encode a protein of 439 amino acids. The chromosomal localization of this human sequence was further determined by the Radiation Hybrid Mapping method (Genethon/Evry) using, for PCR screening, oligonucleotides designed from the sequence of an intronic region. The corresponding transcription unit was localized on chromosome 4, close to the D4S1576-D4S1579 interval (LOD scores of 15.56 and 16.06, respectively). Interestingly, this locus is synthetic with (i.e., contains a set of genes conserved in) the locus containing the murine CCR4 gene (chromosome 3, B–D region; see Ref. 12).

Amino Acid Conservation among yCCR4-related Proteins—
The predicted amino acid coding regions of the human, murine, Xenopus, and yeast proteins were compared, using the J. HEIN Multiple Sequence Alignment Program (Fig. 2). The Xenopus, murine, and human proteins appear to lack the N-terminal region of yeast CCR4 (from amino acids 1 to 432), which contains a leucine-rich repeat region (amino acids 350–467) and two activation domains (6). Yet, they display a significant similarity (close to 30%, including conservative amino acid changes) with the yeast CCR4 protein within their C-terminal domains, in a region corresponding to the second and third exons of the Xenopus, murine, and human genes. High sequence similarity (76% similarity) is also observed between the mammalian and Xenopus proteins but again severely restricted to the region corresponding to the second and third exons of the genes, the coding region of the mammalian CCR4 first exons being significantly longer (62/64 versus 22 amino acids) than, and divergent from, that of Xenopus. Finally, the murine and human CCR4 proteins display strong similarity (93%) equally distributed over the three exons. This similarity stops immediately upstream of their respective initiation codon, suggesting that these codons are those actually used for initiation of translation.

A search for putative amino acid motifs in the sequence of mCCR4 (using MOTIF finder) revealed several phosphorylation and glycosylation sites. The leucine-rich repeat region...
present in the yCCR4 protein at position 350–467 (5) could not be found in the amino acid sequence of the mammalian proteins. Moreover, although a consensus leucine zipper-like motif was identified close to the N-terminal region of the putative nocturnin protein (from 64 to 91; Ref. 10), no such motif could be found in the human and mouse proteins, in particular because of the absence of a leucine at position 125 and 123, respectively (indicated with a star in Fig. 2). Interestingly, search against a database of conserved blocks (19) revealed an 8-amino acid homology (PDILCLQEV, amino acids 186–194 in mCCR4) with a conserved motif involved in Mg\(^{2+}\) or Mn\(^{2+}\) binding among the AP- (Apurinic/Apyrimidinic) endonuclease family I (20). This motif is one of the most conserved from yeast to human CCR4.

**Genomic Organization of the Murine CCR4 Gene**—The genomic organization of the murine mCCR4 gene was assessed by different methods. Introns and locations of intron/exon boundaries were identified by PCR reactions using a series of primers derived from the sequence of the regions corresponding to the characterized cDNA. To explore the 5′ genomic region of the gene, a 129/sv mouse genomic library (constructed with Sau3A1-restricted DNA) was screened by a PCR-based method (see "Experimental Procedures"), and the unique phage clone obtained was directly sequenced. It contained part of intron 1 and 430 nt of exon 1, up to a Sau3A site in exon 1. More 5′-domains were characterized by inverted PCR using genomic DNA, first restricted and then self-ligated, and divergent primers derived from the newly identified sequences. As illustrated in Fig. 3A, the mouse mCCR4 gene spans almost 27 kb of genomic DNA, with three exons and two introns (as previously observed, interestingly, for the Xenopus nocturnin gene, see Ref. 10). Exon 1 contains 185 bp of coding sequence and 836 bp of 3′-untranslated region. In mouse genomic DNA, these two exons are separated by 1.8 kb of intron 2 sequence.

Surprisingly, genomic analysis of intron 1 revealed that the IAP insertion is not common to all laboratory mouse strains. In

**Fig. 2.** Comparison of the predicted amino acid sequence of the human (hCCR4) and the murine (mCCR4) CCR4-related proteins, of the Xenopus nocturnin, and of the yeast CCR4 factor (yCCR4). Sequences are represented with the single-letter code, with numbers referring to amino acid positions for each protein (GenBank accession numbers U74761 and S50459 for nocturnin and yCCR4, respectively; and AF183960 and AF183961 for mCCR4 and hCCR4, respectively). The first 21 N-terminal amino acids from nocturnin, as well as the 431 N-terminal amino acids from yCCR4, could not be aligned with the mammalian CCR4-related proteins. Dashes represent gaps introduced to optimize alignment. Amino acid identities of the human CCR4 protein with at least one of the three other proteins are boxed. Positions corresponding to exon boundaries in mCCR4 are indicated. The asterisk is positioned above the leucine residue of the nocturnin leucine zipper-like motif not found in the murine and human proteins.
a previous paper, we had reported that the IAP-AR transcripts observed in the liver of aged mice were induced in the four inbred strains tested, i.e. C57BL/6, C57BL/10, BALB/c, and DBA/2. We have now characterized eight inbred laboratory strains as well as two strains of wild-type mice (PKW and SEG). The position of the IAP insertion in the first intron of the gene is shown. Arrows indicate the transcription start sites. Positions of the BglII restriction sites (B) and of the probes (5' and 3') used for gene mapping are indicated. B, distribution of inbred laboratory and wild-type mouse strains according to the IAP insertion within the mCCR4 gene. Presence of the IAP insertion was determined by Southern blot analysis upon restriction of the DNA from each mouse strain with BglII and hybridization with the probes indicated in panel A. Lack of IAP insertion was confirmed by PCR using primers bracketing the integration site. C, characterization of the IAP insertion site: nucleotide sequence of the IAP 5'- and 3'-ends and cellular flanking DNA in BALB/c mice (top) and of the target site without IAP insertion in CBA mice (bottom). IAP nucleotides are in uppercase letters, cellular DNA sequence is in lowercase letters, and the 6-bp target site duplication is delineated.

Identification of a CpG Island at the 5'-End of the mCCR4 Gene and Characterization of the mCCR4 Promoter—The transcription start sites of the murine mCCR4 gene were determined using 5'-RACE PCR with reverse primers in exon 1 (Fig. 1B). Two major bands were amplified and cloned. For each band, the 5'-end of several products was sequenced and found identical, ending at the same positions (Fig. 4A): the smaller product revealed a first start site, defined hereafter as position 1, located 110 bp upstream of the AUG codon and most probably responsible for the major 2.7 kb transcript; and the other band revealed a second start site, 400 bp upstream, possibly associated with the fainter 3.1-kb transcript. Both transcription start sites are included within a nucleotide sequence that fits moderately with the transcription start site consensus YYAN(1)/TATA/YY (21, 22).

Analysis of the genomic sequences encompassing the 5'-end of the murine mCCR4 gene discloses a CpG island, a marker of promoter regions of the so-called "housekeeping," or ubiquitously expressed, genes (although some tissue-specific genes also have these characteristic features, Ref. 23). This region extends from nt -413 to +525 (Fig. 4B), has a high GC content (% G + C = 0.72) and an observed overexpected CpG/GpC ratio of 1.1 (mean genome value 0.2; see Ref. 23). Moreover, screening of the -540 to +400 region for putative transcription factor binding sites (using the TRANSFAC program), revealed features common to CpG island-containing promoters (23): two GC boxes, which are potential Sp1 binding sites, can be identified 110 bp upstream of the major transcription start site (Fig. 4A), whereas neither a TATA- nor a CAAT-box motif is present. Interestingly, a putative binding site for NF-kB can be found at position -220.

To determine whether the 5'-end of the mCCR4 gene possesses a promoter activity and to identify important regulatory domains, 5'-deletion fragments of this region (with a fixed 3'-end at position +114) were generated through restriction

**FIG. 3.** Genomic organization of the mCCR4 gene in common laboratory mouse strains. A, genomic organization of the IAP-tagged mCCR4 gene in the prototype DBA/2 mouse and of the IAP-free mCCR4 gene in the CBA mouse. The exon-intron sequences are drawn to scale, with the coding and untranslated regions represented with black and gray boxes, respectively. The position of the IAP insertion in the first intron of the gene is shown. Arrows indicate the transcription start sites. Positions of the BglII restriction sites (B) and of the probes (5' and 3') used for gene mapping are indicated. B, distribution of inbred laboratory and wild-type mouse strains according to the IAP insertion within the mCCR4 gene. Presence of the IAP insertion was determined by Southern blot analysis upon restriction of the DNA from each mouse strain with BglII and hybridization with the probes indicated in panel A. Lack of IAP insertion was confirmed by PCR using primers bracketing the integration site. C, characterization of the IAP insertion site: nucleotide sequence of the IAP 5'- and 3'-ends and cellular flanking DNA in BALB/c mice (top) and of the target site without IAP insertion in CBA mice (bottom). IAP nucleotides are in uppercase letters, cellular DNA sequence is in lowercase letters, and the 6-bp target site duplication is delineated.

**FIG. 4.** Characteristic features of the 5'-domain of the mCCR4 gene. A, nucleotide sequence of the mCCR4 promoter region (GenBank™ accession number AF183962). The two transcription start sites (arrows) and the corresponding transcription initiator sequences (init) are indicated. The ATG start codon is boxed. The sequences of potential binding sites for the Sp1 and NF-kB factors are underlined. B, distribution of CpG and GpC dinucleotides in the 5' genomic region of mCCR4. The indicated CpG island is defined according to the criteria in Ref. 23 as a region with a percentage of G + C >50 and a CpG/GpC ratio >0.6. Exon 1 with the untranslated (shaded) and translated (black) regions is boxed, non-exonic regions are drawn as a line, and the two transcription start sites are positioned with arrows.
endonuclease digestion and cloned in the sense or antisense orientation upstream of a luciferase reporter gene (Fig. 5A). These constructs, as well as a promoterless luciferase vector and an SV40 promoter-enhancer luciferase vector as controls, were introduced by transfection into 3T3 cells (which express the endogenous mCCR4 gene). As shown in Fig. 5B, the 600-bp fragment spanning nt -500 to +114 of the putative promoter induces a 200-fold increase in luciferase activity as compared with the control promoterless basic vector. Deletion to nt -227 (thus eliminating the first initiator signal) results in only a 2-fold reduction in promoter activity. Deletion to nt -100, removing the putative Sp1 and NF-xB sites, results in a dramatic (more than 30-fold) reduction in promoter activity. Surprisingly, further deletion to position -30 restored the transcriptional activity up to 50-fold that of the promoterless control vector, suggesting that a repressive element might reside in this region. Finally, further deletion to position -4 dramatically abolished promoter activity, most probably as a result of the deletion of the initiator site. This analysis therefore provides evidence that the region between nt -500 and +4, encompassing the two transcriptional initiation sites, possesses elements with positive and negative effects on transcription, and that the most proximal sequences, between nt -30 and +114 are sufficient for promoter activity. Several CpG island-containing genes have promoters that can initiate transcription in both orientations (24–26). The same fragments as above were therefore tested for their ability to promote transcription in the reverse orientation. As shown in Fig. 5B, all fragments tested initiate transcription almost as efficiently as the corresponding fragments in the sense orientation, thus clearly demonstrating that the mCCR4 promoter can direct initiation bidirectionally. Interestingly, this transcriptional activity in reverse orientation requires the minimal +114/+4 domain, as evidenced by the ability of this fragment to promote an 80-fold increase in luciferase activity as compared with the promoterless control vector. This fragment initiates transcription efficiently in the reverse and not in the normal orientation, most probably because the initiator element, which is deleted in this fragment and is essential for transcription in the normal orientation, is not part of the reverse promoter.

**In Vivo Expression of the Murine mCCR4 Gene—**To get insight into the putative function of the mCCR4 gene, we analyzed its pattern of expression in several mouse organs. For each organ, mice of a strain with (DBA/2) or without (CBA) the IAP insertion were compared. Total RNAs were extracted and analyzed by Northern blot, using as a probe a 1-kb fragment encompassing exons 2 and 3 of the mCCR4 gene (Fig. 1). As illustrated in Fig. 6, and as expected for a “housekeeping” gene, mCCR4 transcripts are observed in all the organs tested, although at a variable level. Interestingly, comparison of this pattern in both strains (the one with the IAP insertion in intron 1 and the one without) reveals no difference, neither in the size nor in the relative abundance of the transcripts from one organ to another, therefore suggesting that the insertion of the IAP is neutral for transcription of the mCCR4 gene. One major exception, as previously reported, concerns transcription in the liver of aged mice where IAP LTR-initiated chimeric transcripts extending through the mCCR4 exons 2 and 3 are induced at a high level. Finally, it is worth mentioning that expression of mCCR4 is high in the retina, as well as in the brain, both organs known to be involved in the circadian clock in mammals (27, 28).

**DISCUSSION**

A Protein with Similarities to the Yeast CCR4 General Transcription Factor, Conserved among Vertebrates—In this report, we have identified, for the first time in mammalian genes related to the yeast gene encoding the yCCR4 general transcription factor. The murine and human genes are located on synaptic chromosomal loci and encode putative proteins disclosing high levels of identity. Furthermore, the murine and human yCCR4-related gene products are closely related to a *Xenopus* yCCR4-related protein, named nocturnin (10). The region of similarity corresponds to the almost complete sequence of the vertebrate proteins, but only to the C-terminal half of the yeast factor, which is twice as long as the vertebrate proteins. This region was previously demonstrated to be necessary but not sufficient alone for transcriptional activation by yCCR4 (6), but its exact function in yeast has not been elucidated. The two N-terminal activation domains and the leucine-rich repeats described in yCCR4 are absent from the vertebrate proteins. No other such leucine-rich tandem repeats could be found in the amino acid sequence of the mammalian proteins. Similarly, the consensus leucine zipper-like motif described near the N-terminal region of the nocturnin protein, from amino acids 64–91 (10), is not found in the human and mouse proteins. Although the presence

---

3 W. Barbot, unpublished data.
of a zipper motif in nocturnin but not in the murine and human yCCR4-related proteins is conceivable, the occurrence of a proline residue, incompatible with an α-helical secondary structure, in the zipper region of nocturnin (as noted by the authors themselves) casts some doubt on the existence of such a functional motif in nocturnin as well. Finally, based on a previous characterization of the 3′-end of the murine yCCR4-related gene by both Southern blot analysis and chromosome in situ hybridization (12), it is very likely that at least the murine gene is a single copy gene. Accordingly, the observed difference in the length of the vertebrate CCR4-related proteins and the yeast CCR4 factor most probably reflects an evolution from a multiple domain transcription factor of large size, as observed for yCCR4, to the smaller vertebrate CCR4-related factors, which only contain part of the functional domains of yCCR4. It is conceivable that the N-terminal part of yCCR4, not found in the CCR4-related vertebrate proteins, participates in a multicomponent vertebrate transcription complex as a separate protein element. Such an evolutionary “splitting” of the yCCR4 factor might have favored more complex regulations to take place within the CCR4 transcription complex.

Possible Functions of the Murine yCCR4-related Gene—Analyses of the sequence and pattern of expression of mCCR4 have revealed characteristic features of most “housekeeping” genes: the mCCR4 promoter contains a CpG island (high G/C content and CpG/GpC value, Sp1 binding sites, absence of a TATA box), and the gene is ubiquitously expressed in most mouse tissues. As such, mCCR4 is likely to be involved in essential cellular processes. Deletion analyses have revealed that the putative Sp1 and NF-κB binding sites within the promoter sequence are important for gene activity as well as other domains, although not containing easily recognizable binding sites, which disclose either positive or negative effects. Yet, the physiological relevance of these cis-regulatory sequences remains to be demonstrated in vivo. Finally, we have shown that the mCCR4 promoter directs transcription in both orientations. This might suggest the existence of an upstream divergent gene sharing a bidirectional promoter with the mCCR4 gene, a feature common to other CpG island-containing housekeeping genes (24–26).

The nocturnin yCCR4-related gene was previously demonstrated to exhibit rhythmic expression in Xenopus retina, and therefore to function as a clock-controlled gene or a component of the clock itself (10). Although we have not been able to obtain evidence for such a circadian-regulated expression of mCCR4 in the murine retina, it remains plausible that mCCR4 is involved in a circadian regulation. Actually, in mammals the circadian clock resides primarily in the suprachiasmatic nucleus of the hypothalamus and not in the retina (Refs. 27 and 29; but see also Ref. 28), and it is also known that clock orthologs within different species share extensive structural features but are subjected to different regulatory schemes. For instance, the Drosophila clock and tim transcripts show daily rhythm, whereas their murine orthologs do not oscillate (30–33). Accordingly, mCCR4 which is highly expressed both in the retina and the brain could have a function in the circadian clock, as shown for nocturnin, but not be rhythmically regulated. It is also worth mentioning that many clock-regulated genes are housekeeping genes.

Consequences of IAP Insertion in the Intron of the mCCR4 Gene—There are several examples where IAP sequences have disrupted gene expression as a consequence of their insertion into an intron (34–36), either by interfering with transcription or by destabilizing the message. In addition, aberrant chimeric transcripts from a targeted gene, involving splice sites within IAP sequences, are frequently observed (37). Rather unexpectedly, the insertion of an IAP sequence within the first intron of the mCCR4 gene appears to be neutral with regard to transcription of mCCR4. Indeed, sequencing of the two mCCR4 transcripts revealed no alternative splicing involving some IAP site, and Northern blot analyses using mouse strains with or without IAP insertion have failed to detect any differences in the size of the transcripts (that could have suggested exon skipping and/or transcription termination) and in the pattern of expression of mCCR4 between the two mouse strains. Transcriptional read-through previously described for the LTR-initiated IAP-CCR4 chimeric transcripts in the liver of old mice has led us to postulate the existence of mechanisms abolishing the polyadenylation efficiency of the 3′-LTR of the inserted IAP element (12). It is likely that the same mechanism accounts for the absence of transcription termination within the IAP 3′-LTR of the mCCR4-promoted transcripts in all organs tested. The molecular mechanisms responsible for the occurrence of this read-through might involve, as observed in Drosophila for genes mutated by retrotransposon insertion (e.g. w spriclot mutation, Ref. 38), suppressors of transcription termination, which still remain to be identified in mammals. Whatever the underlying molecular mechanisms, the strong selective pressure that preserved mCCR4 from the deleterious effect of IAP insertion suggests an important role for this factor in mammals. Transgenesis experiments with a mutated gene most probably will be necessary to unravel the role of mCCR4 in the mouse, as well as genetic analyses of the corresponding chromosomal loci in humans.

Acknowledgments—We thank G. Hamard (U129, Paris) for providing us with the 129sv mouse genomic library, J. L. Guenet (Institut Pasteur, Paris) for the PWK and SEG mice, and Y. Courtois and J. C. Jeanny (U450, Paris) for help with the retina experiments. J. Weissenbach and G. Gyapay (Genethon, Evry) are especially acknowledged for the localization of the human CCR4-related gene, and L. Corbo and D. Prévôt (U453, Lyon) are acknowledged for helpful discussions and suggestions and for communication of unpublished results.
REFERENCES
1. Denis, C. L. (1984) Genetics 108, 833–843
2. Denis, C. L., and Malvar, T. (1990) Genetics 124, 283–291
3. Liu, H.-Y., Toon, J. H., Chiang, Y.-C., Draper, M. P., Johnson, L. H., and Denis, C. L. (1997) EMBO J. 16, 5289–5298
4. Liu, H.-Y., Badarinarayana, V., Audino, D. C., Rappsilber, J., Mann, M., and Denis, C. L. (1998) EMBO J. 17, 1096–1106
5. Malvar, T., Biron, R. W., Kaback, D. B., and Denis, C. L. (1992) Genetics 132, 951–962
6. Draper, M. P., Liu, H.-Y., Nelson, A. H., Mosley, S. P., and Denis, C. L. (1994) Mol. Cell. Biol. 14, 4522–4531
7. Verdone, L., Cesari, F., Denis, C. L., Di Mauro, E., and Caserta, M. (1997) J. Biol. Chem. 272, 30828–30834
8. Schild, D. (1995) Genetics 140, 115–127
9. McKenzie, E. A., Kent, N. A., Dowell, S. J., Moreno, F., Bird, L. E., and Mellor, J. (1993) Mol. Gen. Genet. 240, 374–386
10. Green, C. B., and Besharse, J. C. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 14884–14888
11. Draper, M. P., Salvador, C., and Denis, C. L. (1995) Mol. Cell. Biol. 15, 3487–3495
12. Puech, A., Dupressoir, A., Lloreu, M.-P., Mattei, M.-G., and Heidmann, T. (1997) J. Biol. Chem. 272, 5965–6003
13. Kuff, E. L., and Lueders, K. K. (1988) Adv. Cancer Res. 51, 183–276
14. Heidmann, O., and Heidmann, T. (1991) Cell 64, 159–170
15. Kuff, E. L. (1990) Cancer Cells 2, 388–400
16. Dupressoir, A., Puech, A., and Heidmann, T. (1995) Biochim. Biophys. Acta 1264, 397–402
17. Israel, D. I. (1993) Nucleic Acids Res. 21, 2627–2631
18. Dupressoir, A., and Heidmann, T. (1996) Mol. Cell. Biol. 16, 4495–4503
19. Henkoff, S., and Henkoff, J. G. (1994) Genomics 19, 97–107
20. Seki, S., Ikeda, S., Watanabe, S., Hattuscuki, M., Tatsuuki, K., Akiyama, K., and Zhang, H. (1991) Biochim. Biophys. Acta 1079, 57–64
21. Smale, S. T., and Baltimore, D. (1988) Cell 57, 103–113
22. Bucher, P. (1990) J. Mol. Biol. 212, 563–578
23. Gaudrner-Gauden, M., and Frommer, M. (1987) J. Mol. Biol. 196, 261–282
24. Liao, W.-C., Ash, J., and Johnson, L. F. (1994) Nucleic Acids Res. 22, 4044–4049
25. Schilling, L. J., and Farnham, P. J. (1995) Cell Growth Differ. 6, 541–548
26. Qvist, H., Sjöström, H., and Noren, O. (1998) Biol. Chem. 379, 75–81
27. Ralph, M. R., Foster, R. G., Davis, F. C., and Menaker, M. (1990) Science 247, 975–978
28. Tosini, G., and Menaker, M. (1996) Science 272, 419–421
29. Welsh, D. K., Logothetis, D. E., Meister, M., and Reppert, S. M. (1995) Neuron 14, 697–706
30. Sun, Z. S., Albrecht, U., Zhuchenko, O., Bailey, J., Eichele, G., and Lee, C. C. (1997) Cell 90, 1003–1011
31. Bae, K., Lee, C., Sidote, D., Chuang, K., and Edery, I. (1998) Mol. Cell. Biol. 18, 6142–6151
32. Sangoram, A. M., Saez, L., Antoch, M. P., Gekakis, N., Staknis, D., Whiteley, A., Fruechte, E. M., Vitaterna, M. H., Shimomura, K., King, D. P., Yeung, M. W., Weitz, C. J., and Takahashi, J. S. (1998) Neuron 21, 1101–1113
33. Zyka, M. J., Shearman, L. P., Levine, J. D., Jin, X., Weaver, D. R., and Reppert, S. M. (1998) Neuron 21, 1115–1122
34. Xiao, G.-H., Jin, F., and Yeung, R. S. (1995) Oncogene 11, 81–87
35. Hamilton, B. A., Smith, D. J., Mueller, K. L., Kerrebrock, A. W., Bronson, R. T., van Berkel, V., Daly, M. J., Kruglyak, L., Pat Reeve, M., Nemhauser, J. L., Hawkins, T. L., Rubin, E. M., and Lander, E. S. (1997) Nature 382, 711–722
36. Gwynn, B., Lueders, K., Sands, M. S., and Birkenmeier, E. H. (1998) Mol. Cell. Biol. 18, 6474–6481
37. Vasicek, T. J., Zeng, L., Guan, X.-J., Zhang, T., Costantini, F., and Tilghman, S. M. (1997) Genetics 147, 777–786
38. Zachar, E., Davison, D., Garza, D., and Bingham, P. M. (1985) Genetics 111, 495–515