

Research article

## Identification of four families of yCCR4- and Mg<sup>2+</sup>-dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAFI/hPOP2 binding

Anne Dupressoir<sup>1</sup>, Anne-Pierre Morel<sup>2</sup>, Willy Barbot<sup>1</sup>, Marie-Paule Loireau<sup>1</sup>, Laura Corbo<sup>2</sup> and Thierry Heidmann\*<sup>1</sup>

Address: <sup>1</sup>Unité des Rétrovirus Endogènes et Éléments Rétroïdes des Eucaryotes Supérieurs, UMR 1573 CNRS, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France and <sup>2</sup>Biologie des Gènes Suppresseurs de Tumeur, INSERM U453, Centre Léon Bérard, 28 rue Laënnec, 69373 Lyon Cedex 08, France

E-mail: Anne Dupressoir - dupresso@igr.fr; Anne-Pierre Morel - morela@lyon.fnclcc.fr; Willy Barbot - wbarbot@igr.fr; Marie-Paule Loireau - mloireau@igr.fr; Laura Corbo - corbo@lyon.fnclcc.fr; Thierry Heidmann\* - heidmann@igr.fr

\*Corresponding author

Published: 22 November 2001

Received: 9 August 2001

BMC Genomics 2001, 2:9

Accepted: 22 November 2001

This article is available from: <http://www.biomedcentral.com/1471-2164/2/9>

© 2001 Dupressoir et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved along with the article's original URL. For commercial use, contact [info@biomedcentral.com](mailto:info@biomedcentral.com)

### Abstract

**Background:** The yeast yCCR4 factor belongs to the CCR4-NOT transcriptional regulatory complex, in which it interacts, through its leucine-rich repeat (LRR) motif with yPOP2. Recently, yCCR4 was shown to be a component of the major cytoplasmic mRNA deadenylase complex, and to contain a fold related to the Mg<sup>2+</sup>-dependent endonuclease core.

**Results:** Here, we report the identification of nineteen yCCR4-related proteins in eukaryotes (including yeast, plants and animals), which all contain the yCCR4 endonuclease-like fold, with highly conserved CCR4-specific residues. Phylogenetic and genomic analyses show that they form four distinct families, one of which contains the yCCR4 orthologs. The orthologs in animals possess a leucine-rich repeat domain. We show, using two-hybrid and far-Western assays, that the human member binds to the human yPOP2 homologs, i.e. hCAFI and hPOP2, in a LRR-dependent manner.

**Conclusions:** We have identified the mammalian orthologs of yCCR4 and have shown that the human member binds to the human yPOP2 homologs, thus strongly suggesting conservation of the CCR4-NOT complex from yeast to human. All members of the four identified yCCR4-related protein families show striking conservation of the endonuclease-like catalytic motifs of the yCCR4 C-terminal domain and therefore constitute a new family of potential deadenylases in mammals.

### Background

The Carbon Catabolite Repressor 4 factor in *Saccharomyces cerevisiae* (yCCR4) regulates the expression of a number of genes involved in nonfermentative growth [1], in cell wall integrity [2], in UV sensitivity [3], and in me-

thionine biosynthesis [4]. yCCR4 is an essential component of several complexes involved in transcription. A first one is the CCR4-NOT multi-subunit group of proteins, which comprises at least two complexes of 1.0 and 1.9 MDa. The 1.0 MDa complex contains yCCR4, yPOP2

(also referred to as yCAF1) and at least five yNOT1-5 proteins [5-7]. In this complex, yPOP2 independently binds to yCCR4 and yNOT1, and is absolutely required for yCCR4 to associate with the 1.0 MDa complex [6,7]. In the 1.9 MDa complex, yCCR4 binds to proteins such as yDBF2, a cell-cycle regulated protein kinase [2], yCAF4 and yCAF16, and is essential for the interaction of both yCAF4 and yCAF16 with ySRB9, a component of the RNA polIII holoenzyme [8]. yCCR4 was also reported to be associated with Paf1, Cdc73, and Hpr1 in a RNA polIII complex distinct from the SRBP-containing holoenzyme [9]. Accordingly, yCCR4 contains, in its central region, a leucine-rich repeat (LRR) domain [10] which was demonstrated to be necessary for yCCR4 binding to the yPOP2 [5,11], yDBF2 [2], yCAF4 and yCAF16 [8] components of the CCR4-NOT complex. Moreover, its N-terminus discloses two activation domains which are required for transcriptional activation [10]. Hypotheses concerning yCCR4 function in yeast have been documented recently. First, Dlakic [12] and Hofmann *et al* [13] showed that the yCCR4 C-terminus contains a fold related to the Mg<sup>2+</sup>-dependent endonuclease core, suggesting that it may function as a nuclease. Second, yCCR4 was shown to be a component, in association with yPOP2, of the major yeast cytoplasmic mRNA deadenylase complex, and to be required for efficient poly(A)-specific mRNA degradation [14,15]. yCCR4 might be a catalytic subunit of this complex, because of its nuclease-like domain. Hence, yCCR4-associated complexes are likely to fulfill fundamental functions in gene regulation, and it is therefore of general interest to determine if similar complexes exist in mammals. Several components of the yCCR4-associated complexes have already been identified in humans: two homologs of yPOP2 (named hCAF1 and hPOP2/hCALIF) [16-19], and homologs of NOT proteins (named hNOT1, hNOT2, hNOT3, hNOT4) [19]. Moreover, interactions between these proteins were demonstrated [19,20]. Yet, at the present time there has been no report of a human protein that would be structurally and functionally close to yCCR4. We [21] and others [22] previously characterized genes for highly related proteins in *X. laevis*, *M. musculus* and *H. sapiens* (named *nocturnin*, "*mCCR4*" and "*hCCR4*"), disclosing circadian expression [22], and with C-terminus displaying significant similarity with the C-terminus of yCCR4 (close to 30%). Yet, these proteins lack the N-terminal region of yCCR4 (i.e. aa 1 to 505), which contains the two activation domains and the leucine-rich repeat region necessary for protein-protein interactions. To identify orthologs of yCCR4 in higher eukaryotes and unravel the phylogenetical relationships between yCCR4 and the previously identified genes, we made a systematic search for other proteins related to yCCR4. Here we report the identification of nineteen yCCR4-like proteins in eukaryotes and show, by phylogenetic and genomic analyses,

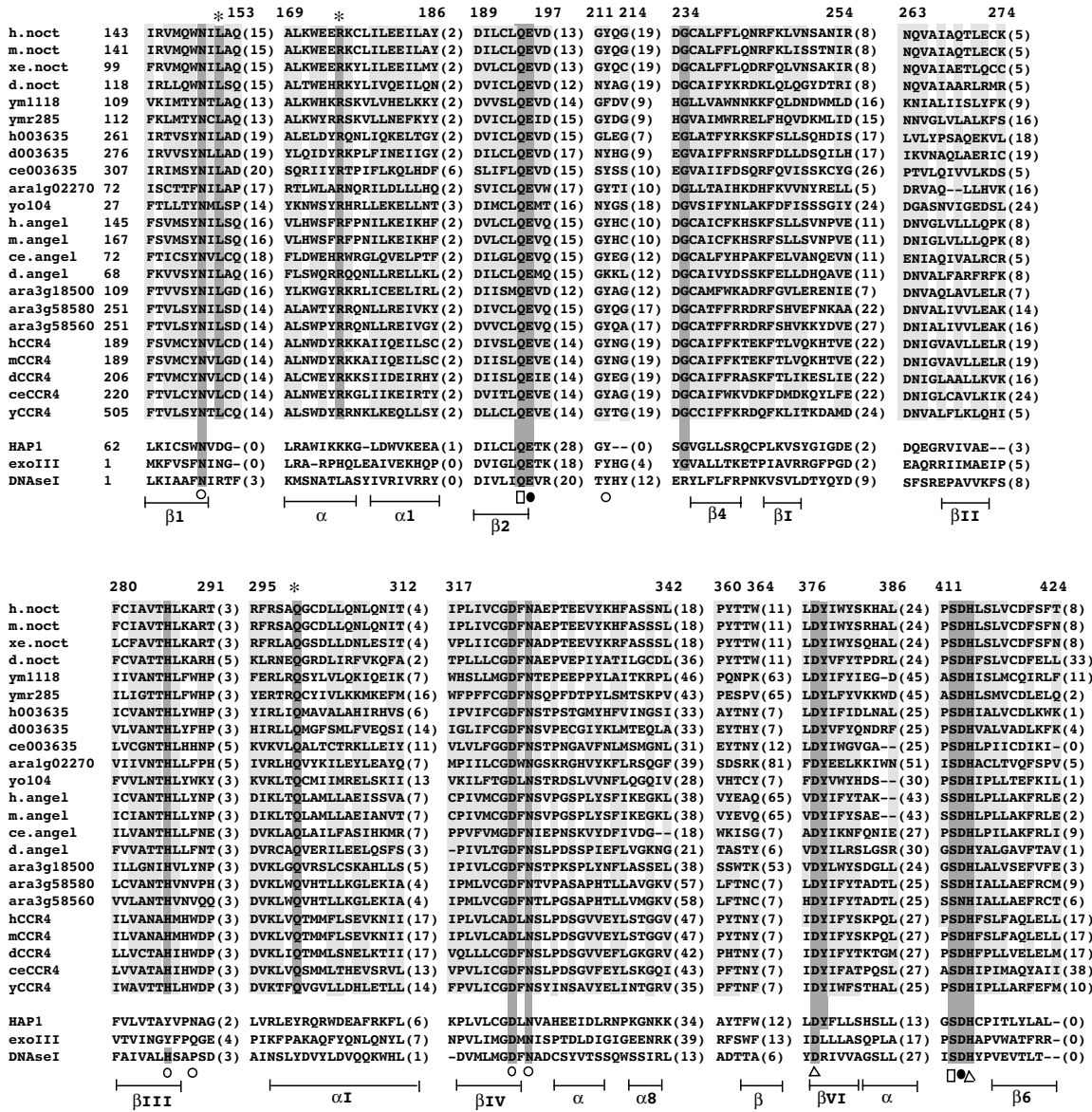
that they are grouped into four distinct families, one of which contains the yCCR4 orthologs: these orthologs, in animals, have conserved the yCCR4 leucine-rich repeat, and we show, using two-hybrid assays and far-Western experiments, that the human protein binds to the human yPOP2 homologs, i.e. hCAF1 and hPOP2, in a LRR-dependent manner. Amino acid alignments show that the endonuclease-like catalytic motifs of the yCCR4 C-terminal domain are strictly conserved among all yCCR4-related proteins and identify CCR4-specific residues in this domain. The results therefore strongly suggest the existence of a conserved CCR4-NOT complex in human cells, and of a new family of potential deadenylases in mammals.

## Results

### Characterization and phylogeny of yCCR4-related proteins

Sorting out of yCCR4-related proteins was performed by i) searches in databases using the yCCR4 C-terminal domain (aa 505-837) conserved in the three previously identified yCCR4-like proteins [21,22], and ii) PCR-amplification of cDNA libraries or reverse-transcribed RNAs using appropriate primers followed by sequencing, when required (Table 1). The majority of matching hits obtained through BLAST search in eukaryotic databases belongs to *Arabidopsis thaliana* (plants), *Saccharomyces cerevisiae* (fungi), and *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* (animals/metazoa), as expected from the abundance of these sequences in present-day databases. Sequences obtained from bacteria databases were not included in the phylogenetic analysis because of their poor lod-score, but were considered in subsequent analyses. After compilation of the positive hits and, for several of them, PCR amplification of the corresponding cDNAs and sequencing, the open reading frame of each of the 19 matching proteins was determined (Fig. 1B, Table 1). A comparative analysis of the full-length sequence of all yCCR4-related proteins indicates that they all share, as expected from the search strategy, a highly conserved yCCR4-related C-terminus (hatched box in Fig. 1B), but, conversely, their N-terminal regions are highly divergent (except for the LRR-containing group of proteins, see below). A multiple sequence alignment performed with the C-terminal regions (see Fig. 2) allowed the establishment of phylogenetic relationships among all proteins (see Material and Methods). As illustrated in Fig. 1A, the tree constructed using the neighbor-joining method (similar results were obtained using the parsimony method, not shown), indeed shows that the proteins can be classified into four major groups, with significant bootstrap values, which can therefore be considered as distinct protein families.





**Figure 2**

Multiple amino acid sequence alignment of the C-terminal region of 23 yCCR4-related proteins, extended to three Mg<sup>2+</sup>-dependent endonucleases, i.e. HAPI (human; accession number M80261), exoIII (E. coli; accession number AAC74819) and DNaseI (bovine; accession number M60606). Only blocks containing more than three positions where residues display at least 60% similarity (calculated from a distance BLOSUM62 matrix, SQX program, Infobiogen) are shown, with the remaining of the sequence indicated by residue numbers in parenthesis. Positions with strict identity are shaded dark grey, and those with >60% similarity light grey. For the other aligned proteins (HAPI, exoIII, and DNaseI), residues identical to those strictly conserved in the yCCR4-related proteins are also shaded dark grey. Residues conserved in the yCCR4-related proteins but not found in the presently aligned endonucleases or other Mg<sup>2+</sup>-dependent endonuclease-fold containing proteins (not shown) are marked with an asterisk. The secondary structure elements (α-helix and β-strand) for HAPI and/or exoIII (see Fig. 4 in [31] and Fig. 1 in [30]) are indicated under the alignment. Important residues involved in the Mg<sup>2+</sup>-dependent endonuclease activity are indicated as follows: empty circles for phosphate binding residues, filled circles for Mg<sup>2+</sup> binding residues, triangles for catalytic residues, and squares for residues involved in orientation and/or stabilization of the catalytic residues. Amino acid positions above the alignment are for h.nocturnin.

**Table 1: yCCR4-related proteins**

| Name <sup>a</sup> | Nucleotide/Amino acid sequence <sup>b</sup> | Sorting procedure (this work) or reference | Corresponding genomic sequence <sup>c</sup> or reference |
|-------------------|---|--|--|
| h.nocturnin       | AAD56548                                    | Dupressoir et al [21]                      | NT-006124  |
| m.nocturnin       | AAD56547                                    | Dupressoir et al [21]                      | Dupressoir et al [21]                                    |
| xe.nocturnin      | AAB39495                                    | Green & Beshare [22]                       | Green & Beshare [22]                                     |
| d.nocturnin       | AY043266                                    | cDNA / PCR                                 | AE003690   |
| ym1118            | NP_013588                                   | -  | NC001145   |
| ymr285            | NP_014012                                   | -  | NC001145   |
| h3635             | file 1                                      | Databases compilation                      | AC016923   |
| d3635             | file 2                                      | Databases compilation                      | AE003635   |
| ce3635            | CAB07271                                    | -  | Z92809   |
| ara1g02270        | At1g02270                                   | -  | NC_003070  |
| yo1042            | NP_014600                                   | -  | NC_001147  |
| h.angel           | file 3                                      | Databases compilation                      | AC025707   |
| m.angel           | file 4                                      | Databases compilation                      | -  |
| ce.angel          | file 5                                      | Databases compilation                      | AF067946   |
| d.angel           | CAA59746                                    | Kurzik-Dumke et al [26]                    | AE003461   |
| ara3g18500        | At3g18500                                   | -  | NC_003074  |
| ara3g58580        | At3g58580                                   | -  | NC_003074  |
| ara3g58560        | At3g58560                                   | -  | NC_003074  |
| hCCR4             | BAA86508                                    | Nagase et al [27]                          | AC025336   |
| mCCR4             | AY043269                                    | CDNA/PCR                                   | -  |
| dCCR4             | AY043267                                    | cDNA/PCR                                   | AE003746   |
| ceCCR4            | AY043268                                    | cDNA / PCR                                 | Z68753   |
| yCCR4             | AAB24455                                    | Malvar et al [10]                          | S50459   |

<sup>a</sup> prefixes are as follows : h, *Homo sapiens*; m, *Mus musculus*; xe, *Xenopus laevis*; d, *Drosophila melanogaster*; ce, *Caenorhabditis elegans*; y, *Saccharomyces cerevisiae*; ara, *Arabidopsis thaliana*.<sup>b</sup>accession numbers (including those of cDNA coding regions sequenced in this work) or file numbers (see Additional Material) for the in silico determined sequences (this work). <sup>c</sup> accession numbers

The formerly identified "mCCR4", "hCCR4" [21] and nocturnin [22] proteins are grouped within the same family (called hereafter the nocturnin family, with the "mCCR4" and "hCCR4" proteins re-named m.nocturnin and h.nocturnin, and the *xenopus* nocturnin protein, xe.nocturnin). BLAST searches identified a putative *drosophila* ortholog from genomic scaffold AE003635. PCR amplification of cDNA *drosophila* libraries, using primers (indicated in Fig. 1C) designed from the predicted open reading frame, allowed us to provide the entire coding sequence of this gene (named d.nocturnin; assigned GenBank accession number AY043266; Table 1). Human, murine, *drosophila* and *xenopus* nocturnin proteins display significant levels of similarity in their C-terminal region (50–95% among themselves and 24–27% with the yCCR4 C-terminus) (Fig. 1B), but not in their N-terminal part. No ortholog of the nocturnin genes could be found in the *C. elegans* genome (although completely sequenced), thus strongly suggesting that there is no such gene in this species. Besides proteins from animals, two yeast predicted proteins (ym1118 and ymr285. Table 1) with significant similarity with the yCCR4 C-terminus (23 and 24%, respectively), also fall within the nocturnin family. Finally, two *arabidopsis*

predicted proteins (Atlg31500 and Atlg31530) were found to be close to the nocturnin family members (not shown), but lacked the most 5' part of the C-terminal region, and were therefore excluded.

The second family, hereafter named the 3635 family, contains as yet unknown proteins from *C. elegans*, *D. melanogaster* and *H. sapiens* (Table 1). The *caenorhabditis* protein (ce3635) corresponds to the CAB07271 predicted protein. The *drosophila* protein (d3635) corresponds to a putative protein that we predicted from genomic scaffold AE003635 (nt 118477 to 120285; see Additional File 1). The human protein (h3635) was predicted from the human genomic sequence AC016923, the human ESTs AI369813, AW967801, AW968624, AW965496, H06019, and by comparison with the murine EST BE634658 for determination of the putative initiation codon (see Additional File 2). It should be noted that although some mouse ESTs do match in BLAST searches, suggesting the existence of a murine 3635 protein, its C-terminal region could not be predicted in its entirety because of the lack of corresponding overlapping ESTs or genomic sequences. Again, 3635 proteins display a conserved C-terminal region (32–41% similar-

| consensus     | P X X <u>a</u> X X <u>a</u> X X L X X L X L s X N X <u>a</u> X X <u>a</u>                                    | position |
|---------------|--|----------|
| <b>hCCR4</b>  | S A S <b>L</b> W S <b>L</b> T H <b>L</b> T A <b>L</b> H <b>L</b> S D <b>N</b> S <b>L</b> S R <b>I</b>        | 44-67    |
| <b>mCCR4</b>  | S S S <b>L</b> W S <b>L</b> T H <b>L</b> T A <b>L</b> H <b>L</b> S D <b>N</b> S <b>L</b> S C <b>I</b>        | 44-67    |
| <b>dCCR4</b>  | S P S <b>L</b> W E F E H <b>L</b> T A <b>L</b> Y <b>L</b> N D <b>N</b> Q <b>L</b> L R <b>L</b>               | 66-89    |
| <b>ceCCR4</b> | S P S <b>L</b> W Q <b>L</b> T H <b>L</b> S A <b>L</b> F <b>L</b> N N <b>N</b> G <b>L</b> T R <b>L</b>        | 80-103   |
| <b>yCCR4</b>  | S A N <b>I</b> F K <b>Y</b> D F <b>L</b> T R <b>L</b> Y <b>L</b> N G <b>N</b> S <b>L</b> T E <b>L</b>        | 350-373  |
| <b>hCCR4</b>  | <b>P</b> S D <b>I</b> A K <b>L</b> H N <b>L</b> V Y <b>L</b> D <b>L</b> S S <b>N</b> K <b>I</b> R S <b>L</b> | 68-89    |
| <b>mCCR4</b>  | <b>P</b> S D <b>I</b> A K <b>L</b> H N <b>L</b> V Y <b>L</b> D <b>L</b> S S <b>N</b> Q <b>I</b> Q S <b>L</b> | 68-89    |
| <b>dCCR4</b>  | <b>P</b> A D <b>V</b> G M <b>L</b> T S <b>L</b> R T <b>L</b> D <b>L</b> S S <b>N</b> K <b>L</b> R S <b>L</b> | 90-112   |
| <b>ceCCR4</b> | <b>P</b> P E <b>I</b> A Q <b>L</b> T N <b>L</b> T M <b>L</b> D <b>I</b> S <b>N</b> N <b>K</b> L R S <b>L</b> | 104-126  |
| <b>yCCR4</b>  | <b>P</b> A E <b>I</b> K N <b>L</b> S N <b>L</b> R V <b>L</b> D <b>L</b> S H <b>N</b> R <b>L</b> T S <b>L</b> | 374-396  |
| <b>hCCR4</b>  | <b>P</b> A E <b>L</b> G N <b>M</b> V S <b>L</b> R E <b>L</b> H <b>L</b> N N <b>N</b> L <b>L</b> R V <b>L</b> | 90-113   |
| <b>mCCR4</b>  | <b>P</b> A E <b>L</b> G N <b>M</b> V S <b>L</b> R E <b>L</b> H <b>L</b> N Y <b>N</b> Q <b>L</b> R V <b>L</b> | 90-113   |
| <b>dCCR4</b>  | <b>P</b> A E <b>L</b> G E <b>L</b> I Q <b>L</b> R E <b>L</b> L <b>N</b> N <b>N</b> F <b>L</b> R V <b>L</b>   | 113-135  |
| <b>ceCCR4</b> | <b>P</b> T E <b>L</b> G D <b>M</b> I S <b>L</b> C H <b>L</b> Y <b>L</b> N N <b>N</b> Q <b>L</b> R V <b>L</b> | 127-149  |
| <b>yCCR4</b>  | <b>P</b> A E <b>L</b> G S <b>C</b> F Q <b>L</b> K Y <b>F</b> Y <b>F</b> F D <b>N</b> M V T T <b>L</b>        | 397-419  |
| <b>hCCR4</b>  | <b>P</b> F E <b>L</b> G K <b>L</b> F Q <b>L</b> Q T <b>L</b> G <b>L</b> K G <b>N</b> P <b>L</b> T Q D I L N  | 114-139  |
| <b>mCCR4</b>  | <b>P</b> F E <b>L</b> G K <b>L</b> F Q <b>L</b> Q T <b>L</b> S <b>L</b> K G <b>N</b> P <b>L</b> T Q D I L N  | 114-139  |
| <b>dCCR4</b>  | <b>P</b> Y E <b>I</b> G K <b>L</b> F H <b>L</b> V I <b>L</b> G <b>L</b> M G <b>N</b> P <b>L</b> Q K E F M N  | 136-161  |
| <b>ceCCR4</b> | <b>P</b> Y E <b>L</b> G K <b>L</b> F R <b>I</b> Q T <b>L</b> G <b>L</b> Q G <b>N</b> P <b>L</b> S P E I S K  | 150-175  |
| <b>yCCR4</b>  | <b>P</b> W E <b>F</b> G N <b>L</b> C N <b>L</b> Q F <b>L</b> G <b>V</b> E G <b>N</b> P <b>L</b> E K Q F L K  | 420-445  |
| <b>hCCR4</b>  | L C L E P D G T R R L L N Y <b>L</b> L D <b>N</b> L S G T A K  | 140-163  |
| <b>mCCR4</b>  | I Y N E P N G T Q K L L T Y <b>M</b> L D <b>N</b> L S F T V N  | 140-163  |
| <b>dCCR4</b>  | I Y H E T N G A Q K I L Q F <b>L</b> L D H L T I N T H   | 162-185  |
| <b>ceCCR4</b> | I L T E K S V T G - L I F Y <b>L</b> R D <b>N</b> R P E I P L  | 176-199  |
| <b>yCCR4</b>  | L Y Q E P D G T R R L L N Y <b>L</b> L D <b>N</b> L S G T A K  | 446-468  |

**Figure 3**

Alignment of the leucine-rich repeats of the human, murine, *drosophila*, *caenorhabditis* and yeast CCR4 family proteins. The consensus shown above is that of the "S.cerevisiae adenylate cyclase" class of 23-amino acid long, proline-containing leucine-rich repeats, which includes the yCCR4 LRR [28]. Conserved or aliphatic residues matching the consensus sequence are boxed or circled, respectively, a denotes an aliphatic residue.

ity among themselves and 25–29% similarity with the yCCR4 C-terminus) but differ in their N-terminal region.

The third family of yCCR4-related proteins (called hereafter the angel family) contains the formerly identified angel *drosophila* gene product, with as yet no identified function ([26], accession number X85743), as well as *M. musculus*, *H. sapiens*, *C. elegans* and *A. thaliana* proteins (Table 1). The murine protein (m.angel) was determined from the assembly of several murine overlapping matching ESTs (AI151868, BF318959, ABO41602, AA647420 ; see Additional File 3). The human protein (h.angel) was predicted from a human full-length cDNA (accession number AL079275) and the corresponding

genomic sequence (accession number AC025707)(see Additional File 4). The *caenorhabditis* protein (ce.angel) derives from the AAC17686 putative protein, predicted from cosmid AF067946, with a modification for a misassignment of a 3' exon (see Additional File 5). The *arabidopsis* protein (ara3g18500) corresponds to the At3g18500 predicted protein. Once more, the human, murine, *drosophila*, *caenorhabditis* and *arabidopsis* angel proteins have a highly conserved C-terminal region (33–88% similarity among themselves and 25–33% similarity with the yCCR4 C-terminus), but differ in their N-terminus.

The fourth family (hereafter named the CCR4 family) contains the formerly identified yeast CCR4 protein and as yet undefined human, murine, *drosophila*, *caenorhabditis* and *arabidopsis* proteins. The human protein (named hCCR4) is putatively encoded by the KIAA1194 full-length cDNA ([27], accession number AB033020). The cDNA encoding the murine protein (named mCCR4) was isolated by us using RT-PCR and primers derived from matching EST clones, and its open reading frame was entirely sequenced (assigned GenBank accession number AY043269; Table 1). Aside their C-terminal regions which display the highest similarity with the yCCR4 C-terminus (35–41%), the N-terminus of the human and murine proteins display a leucine-rich repeat (LRR) region very similar to that of the yeast CCR4 (Fig. 3 and see below). The *drosophila* and *caenorhabditis* proteins were characterized from matching genomic clones (accession number AE003746 and Z68753, respectively) in which we identified a similar LRR domain 20 kb and 4 kb, respectively, upstream of the exons for the yCCR4-like C-terminus. PCR and sequencing using cDNA libraries and primers that mapped upstream of the LRRs and in the C-terminal region (primers indicated in Fig. 1C) demonstrated that the *caenorhabditis* and *drosophila* putative proteins (hereafter named ceCCR4 and dCCR4, with assigned GenBank accession numbers AY043268 and AY043267, respectively; Table 1) also possess, in their N-terminal region, a LRR motif related to that of the yCCR4 protein (Fig. 1B and 3). Alignment of the leucine-rich repeats (Fig. 3) of the human, murine, *drosophila*, *caenorhabditis* and yeast proteins disclosed high conservation of the residues previously shown to define the 23-amino acid-long LRRs, including those of yCCR4 and of the adenylyl cyclase (PxxaxxaxLxxLxLsxNxaxxa, [28]), thus suggesting that the function of this motif has been conserved from yeast to mammals. Besides proteins from animals, two *arabidopsis* predicted proteins (ara3g58580 and ara3g58560. Table 1) also branch, with significant bootstrap values, with the CCR4 family proteins. Interestingly, although their C-terminal regions display significant similarity with the yCCR4 C-terminus (37%), none of these two proteins possesses a LRR domain (see Fig. 1B and Discussion). Finally, it should be noted that the proteins of the CCR4 family are all significantly smaller than the yeast member, all of them lacking the two activation domains (aa 1 to 350) located in the yCCR4 N-terminal region (see Discussion).

Finally, BLAST searches identified a *S. cerevisiae* putative protein, yo1042 (accession number NP\_014600), and an *A. thaliana* predicted protein (ara1g02270; accession number At1g02270) which disclose similarity with the C-terminal part of the yCCR4 protein (29% and 25%, respectively), but which do not cluster in an unam-

biguous manner with any protein from the four families (Fig. 1A). It should be noted that all the *A. thaliana*, *S. cerevisiae*, *D. melanogaster*, *C. elegans* and *H. sapiens* genes disclosing similarities with yCCR4 have most probably been identified, since the genomes of these five species have now been entirely sequenced and since a BLAST search using the N-terminal -instead of the C-terminal- sequence of yCCR4 (aa 1–350) did not reveal any significant similarity with any other protein from the databases.

#### Genomic organization of the yCCR4-related gene families

Phylogenetic relationships between the yCCR4-related proteins was also analysed through the characterization of the genomic organization of the corresponding genes (Fig. 1C). The intron/exon boundaries were determined upon alignment of the coding regions of cDNA sequences with the corresponding genomic sequences available in genomic databases (Table 1). For all families, the genomic organization of several members of the family could be compared. In all cases, the human locus was significantly longer than the *drosophila* and/or *caenorhabditis* locus, as a result of an increase in intron number and/or size. The four families do not behave similarly when considering the evolution of the number of exons, at least for the animal genes. In the CCR4, angel and 3635 families, the *drosophila* genes contain 5, 1 and 1 exons, respectively, whereas the human genes contain 11, 8 and 3 exons, consistent with a gain of introns in evolution. Conversely, a loss of intron can be suspected for the nocturnin family, for which the *drosophila* and mammalian genes disclose 6 and 3 exons, respectively. In all families, yeast genes are devoid of intron and *arabidopsis* genes display the most exons. Although the number of exons is clearly not constant within a given family, intron positions are in several cases conserved -except for the 3635 family, and for *arabidopsis* genes which do not share any intron position with their family members- (see Fig. 1C). In the CCR4 family, 4/4 introns of the *drosophila* gene and 3/7 introns of the *caenorhabditis* gene have positions homologous to those of the human gene. In the nocturnin and angel family, the *drosophila* and human genes share one intron position. Conversely, paralogous genes in a given species (i.e. genes belonging to different families) do not have any intron position in common. These features suggest (at least for the nocturnin, CCR4 and angel families) that proteins which cluster in the same family indeed share a common ancestor, in agreement with the phylogenetic data. They also strongly suggest that introns were independently acquired, in the four families. This could be easily accounted for assuming that the four families were derived by gene duplication, from an "ancestral" gene (or genes) devoid of intron at the time of duplication.

**Conservation, among the *yCCR4*-related proteins, of a  $Mg^{2+}$ -dependent endonuclease-like domain, with *CCR4*-specific residues**

Dlakic [12] and Hofmann *et al*[13] recently reported that the *xenopus* nocturnin and *yCCR4* proteins contain a domain with similarities with the enzymatic core of  $Mg^{2+}$ -dependent endonucleases. Therefore, we investigated whether such sequence similarities also exist for the other *yCCR4*-related proteins with, possibly, conserved features specific for these proteins. The multiple sequence alignment of all *yCCR4*-related C-terminal regions (Fig. 2) allowed the identification of several blocks of local homology, with more than 60% similarity among *yCCR4*-related proteins, as well as residues strictly conserved among these proteins (see below). This alignment could be extended to the enzymatic core of three characteristic  $Mg^{2+}$ -dependent endonucleases: the bovine DNaseI [29] and two apurinic/apyrimidinic (AP) DNA-repair endonucleases, i.e. the *E. coli* Exonuclease III [30] and the human HAP1 protein [31]. These nucleases share the same catalytic residues and form a similar four-layered  $\alpha/\beta$ -sandwich motif. Interestingly, the blocks of homology defined for the *yCCR4*-related proteins all correspond to secondary-structure elements ( $\beta$ -strand or  $\alpha$ -helix) which form the core of AP endonucleases (Fig. 2). In addition, the putative active residues shared by both AP endonucleases and DNaseI [29–32], are strictly conserved in all the *yCCR4*-related proteins: these are residues involved in catalysis (indicated by triangles in Fig. 2),  $Mg^{2+}$  binding (indicated by black circles), orientation or stabilization of catalytic residues (indicated by squares), or interaction with the phosphate group (indicated by empty circles). Hence, all the *yCCR4*-related proteins possess, as observed for *yCCR4*, a C-terminal  $Mg^{2+}$ -dependent endonuclease-like domain. Alignment of the *yCCR4*-related proteins with the  $Mg^{2+}$ -dependent endonucleases discloses another important feature: several residues (Leu 151, Arg 174 and Glu 300, indicated with an asterisk in Fig. 2) can be identified, which are strictly invariant among all the *yCCR4*-related proteins, but have no equivalent in the presently aligned AP-endonucleases and DNaseI, or in other  $Mg^{2+}$ -dependent endonuclease fold-containing proteins previously described (not shown) [12,13]. These *CCR4*-specific amino acids are most probably relevant to functions common to and specific for all *yCCR4*-related proteins (see Discussion).

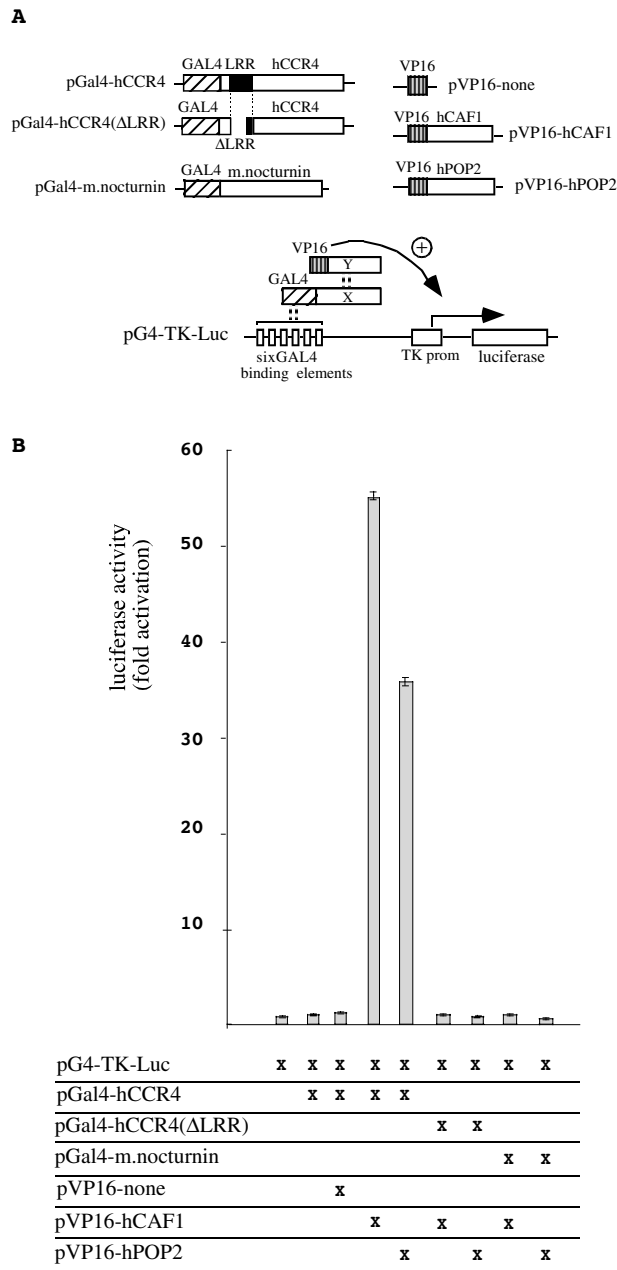
**Interaction of the human homolog of *yCCR4* with hCAF1 and hPOP2/CALIF**

The search for *yCCR4*-related proteins identifies possible orthologs of *yCCR4*, according to the sequence analyses reported above. To confirm these data and gain some hints for the existence of a *CCR4*-NOT complex in mammals, we investigated whether the interaction be-

tween *yCCR4* and *yPOP2* in the *CCR4*-NOT complex is conserved for the human ortholog, h*CCR4*. Two homologs of *yPOP2* have been identified in mammals: the mouse/human CAF1 [11,16,17,33] and, more recently, the human homolog named hPOP2 or hCALIF [18,19]. Both mCAF1 [11] and hPOP2 [19] interact with *yCCR4*, and the *yCCR4* LRR is essential for these interactions. To determine whether *yCCR4*-*yPOP2* interactions are evolutionarily conserved, we examined whether h*CCR4* can bind hCAF1 and hPOP2 in a two-hybrid assay in mammalian cells. The complete h*CCR4* cDNA was obtained (KIAA1194 cDNA, Kazusa DNA Research Institute), and vectors were constructed to express h*CCR4* fused to the DNA binding domain of the yeast GAL4 transcription factor (pGal4-h*CCR4*) (Fig. 4A). A vector expressing hCAF1 (pVP16-hCAF1) or hPOP2 (pVP16-hPOP2) fused to the VP16 activation domain was also produced (Fig. 4A). Two-hybrid protein-protein interaction assays were performed in HeLa cells, with a reporter plasmid (pG4-TK-Luc) containing six GAL4 binding elements upstream of the thymidine kinase (TK) promoter region fused to the luciferase reporter gene (Fig. 4A). In HeLa cells transfected with the fusion constructs and the reporter plasmid, an interaction between GAL4 and VP16 fusion proteins should result in an increase in luciferase expression (Fig. 4A). As observed in Fig. 4B, co-expression of pGal4-h*CCR4* with either pVP16-hCAF1 or pVP16-hPOP2 elicited a significant increase (> 35 fold) in pG4-TK-Luc expression, thus demonstrating association between h*CCR4* and either hCAF1 or hPOP2. This increase is dependent on the presence of the leucine-rich repeats within h*CCR4*, since deletion of four out of the five leucine-rich repeats -within pGal4-h*CCR4*( $\Delta$ LRR), see Materials and Methods- completely abolished the effect (Fig. 4). Finally, a GAL4 fusion vector expressing a murine paralog of h*CCR4*, i. e. m.nocturnin, had no effect when co-expressed with either pVP16-hCAF1 or pVP16-hPOP2 (Fig. 4), thus indicating that m.nocturnin does not interact with hCAF1 or hPOP2, as expected from the absence of an LRR domain in this protein. As a control, pGal4-h*CCR4* alone, or in association with the pVP16-none plasmid, had no effect.

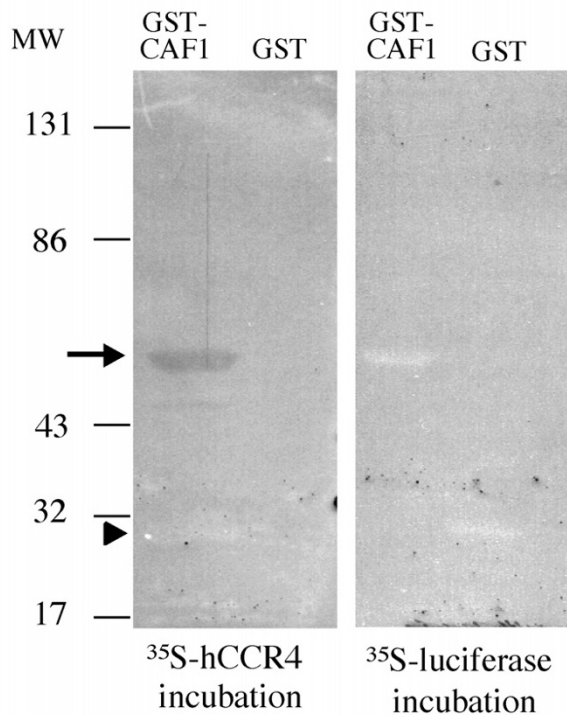
To verify that h*CCR4* can bind directly to hCAF1, we performed a far-Western blot analysis (Fig. 5). Bacterially produced, purified GST-CAF1 and control GST proteins were subjected to SDS-PAGE, transferred onto a membrane and probed with a  $^{35}S$ -methionine-labeled h*CCR4* protein synthesized *in vitro*. As shown in Fig. 5, specific hybridization was observed only with GST-CAF1, and not with GST. As a control, incubation with a  $^{35}S$ -methionine-labeled luciferase protein failed to show any interaction.





**Figure 4**

Two-hybrid assay for in vivo interaction of hCCR4 with hCAF1 and hCALIF/hPOP2. (A) schematic representation of the reporter and fusion protein expression vectors, and rationale of the two-hybrid assay. Coding sequences of hCCR4, hCCR4 deleted for the LRR (hCCR4(ΔLRR)) and m.nocturnin are inserted in-frame with the GAL4 binding domain. The hCAF1 and hPOP2 coding sequences (or no sequence in the pVP16-none vector) are inserted in-frame with the VP16 transactivation domain. The pG4-TK-Luc reporter plasmid contains six GAL4 binding elements, upstream of the thymidine kinase minimal promoter (TK) region fused to the luciferase reporter gene (Luc). An interaction between the GAL4 and VP16 fusion proteins should result in an increase in luciferase expression. (B) hCCR4 interacts with hCAF1 and hPOP2, in a LRR-dependent manner. HeLa cells were cotransfected with a combination of expression vectors, as indicated by crosses. Luciferase activities were measured two days post transfection (see Materials and Methods) and are expressed as the fold increase over the luciferase activity of the reporter vector alone. Bars indicate standard deviation of the mean for at least three independent transfections.



**Figure 5**

hCCR4 interacts with hCAF1 in vitro. Bacterially-expressed and purified GST-hCAF1 and GST (control) proteins were subjected to SDS/PAGE, transferred from the gel onto a polyvinylidene difluoride membrane, and probed with in vitro synthesized  $^{35}\text{S}$ -methionine-labeled hCCR4 (left) or luciferase (control, right) proteins (see Materials and Methods). Autoradiography of the blots demonstrates specific binding of hCCR4 to GST-hCAF1 (57 kDa, arrow), not observed with GST (27 kDa; position of the protein, as determined by a coomassie blue stained gel run in parallel, indicated by an arrowhead). Molecular size markers (in kilodaltons) are indicated. The negative imprints observed in the luciferase control are a common feature associated with reduction of background hybridization to the filter at locations of high protein concentrations.

## Discussion

### Identification of LRR-containing orthologs of yCCR4

In the course of our search for genes encoding proteins displaying significant similarity with the yeast yCCR4 factor, the best hits were found for cDNAs encoding a family of proteins from *C. elegans*, *D. melanogaster*, *M. musculus*, and *H. sapiens*, that we further characterized by both in silico and in vivo approaches. By several criteria, these proteins appear to be yCCR4 orthologs. First, they cluster together with yCCR4 in a phylogenetic tree constructed for the yCCR4-related proteins. Likewise, analysis of their genomic organization revealed that the

*caenorhabditis*, *drosophila* and human CCR4 genes had several introns in common, located at homologous positions, suggesting that they arose from a common ancestor. Second, their N-terminus contain a five leucine-rich repeat domain, disclosing a high level of similarity with that of yCCR4, and with strictly conserved amino acids at consensus positions. Third, the ability of yCCR4 to bind mCAF1 and hPOP2 is conserved for the mammalian protein, since we could demonstrate, using two-hybrid assays and far-Western blot analyses, an in vivo and in vitro interaction between hCCR4 and hCAF1 or hPOP2/CALIF. Moreover we could show, using a deletion within the leucine-rich repeat domain, that this motif is essential for the interaction, a key feature of the previously characterized yCCR4 factor.

Interestingly, the putative hCCR4, mCCR4, dCCR4 and ceCCR4 proteins are much smaller than yCCR4. In fact, they contain the central LRR plus C-terminal region of yCCR4, but they lack its N-terminus (residues 1–350). It is noteworthy that, in yCCR4, these two parts of the protein (1–350 and 350–837) have well-defined and distinct functions, which can be uncoupled. The N-terminal region contains two transactivation domains, one of which is glucose-repressed (residues 1–160). This region can, by itself, activate transcription when targeted to the DNA via a heterologous DNA-binding domain [5]. On the other hand, the LRR plus C-terminal region of yCCR4 cannot mediate transcription activation, but is involved in protein-protein interactions [5]. The yCCR4 LRR is necessary for the binding of several components of the CCR4-NOT complex, including yPOP2 -and mCAF1- [11], yDBF2 [2], yCAF4 and yCAF16 [8], and deletion of the yCCR4 N-terminus does not impair these associations. Consistently, the region encompassing the LRR and part of the C-terminus (residues 302–668) has been defined as the smallest region necessary and sufficient for the association of yCCR4 with yPOP2 [11].

Conservation of the yCCR4 LRR plus C-terminus, together with the absence of its N-terminus in the CCR4 proteins of animals, could be interpreted along two lines. First, the N-terminal domain might mediate yeast-specific functions (e.g. glucose-regulated growth), not conserved in animals because they were unnecessary or redundant in these species. Second, the N- and C-terminal domains observed in the yCCR4 multifunctional protein might have been split into two -or more- independent proteins in animals. If neither interpretation can as yet be favoured, it is noteworthy that yCCR4 evolution parallels that of its partner in the CCR4-NOT complex, namely yPOP2. Indeed, both human homologs of yPOP2, hCAF1 and hPOP2, lack the 148 N-terminal amino acids which are, in yPOP2, required for transcriptional activation, whereas the yPOP2 C-terminus, which

is involved in the interaction with yCCR4, is conserved in hCAF1 [11] and hPOP2 [19]. Similarly, hNOT2, the human homolog of the yNOT2 component of the CCR4-NOT complex, harbors a N-terminal domain divergent from that of yNOT2 -required for transcriptional activation and interaction with the yeast ADA2 factor-, but has conserved the C-terminal domain which is absolutely required for the yCCR4-associated function of yNOT2 [19,20]. Therefore, it looks as if genes of the CCR4-NOT complex have undergone a "concerted" phylogenetic evolution, with the yeast genes encoding multifunctional proteins and the animal genes only encoding "specialized" proteins specific for the CCR4-NOT complex. It is tempting to speculate that partition of the CCR4 or POP2 functions in the course of evolution has been positively selected as resulting in a gain of efficacy. Along this line, Draper *et al*[5] observed that the two transactivation domains present in the yCCR4 N-terminus (residues 1 to 350) are, alone, more potent activators than within the full-length yCCR4 protein. Likewise, mCAF1, which has lost the N-terminal transactivation domain of yPOP2, binds to yCCR4 with a higher affinity than yPOP2 [11].

In yeast, yCCR4 is one of the components of the 1.0 MDa CCR4-NOT complex, which also includes yPOP2 and at least five yNOT proteins. In this complex, yPOP2 independently binds to yCCR4 and to yNOT1. A previous report [19] identified the human orthologs of yPOP2 (hPOP2/CALIF) and yNOT1 (hNOT1), and showed that they have conserved their ability to bind to each other. Hence, the present identification of the human ortholog of yCCR4 and the demonstration that it binds to hPOP2 now provide an additional piece of the puzzle, strongly suggesting the existence of a CCR4-NOT complex in humans. Additionally, yCCR4 interacts, via its LRR domain, with three other proteins (i.e. yDBF2, yCAF4 and yCAF16) of the 1.9 MDa CCR4-NOT complex. Conservation of the LRR domain in hCCR4 suggests that the interaction with these proteins might be conserved as well in humans. Interestingly, we found that hCCR4 also binds to the second human yPOP2 paralog, hCAF1. However, at variance with hPOP2, hCAF1 is not able to interact with hNOT1 (nor with hNOT3) [19]. This suggests that human cells may contain at least two complexes involving a CCR4/POP2 association, a CCR4-NOT-POP2 complex analogous to the CCR4-NOT yeast complex, and a still unknown CCR4-CAF1 containing complex. Interestingly, in yeast only one CCR4-POP2 complex has been described until now, whereas CCR4/POP2 interactions have been found associated with both transcription regulation and mRNA deadenylase activity. This may suggest a "physical" partition of these two CCR4-POP2-associated functions in the course of evolution.

#### **Four families of yCCR4-related genes: function and origin?**

Recently, Dlakic [12] and Hofmann *et al*[13] showed that the yCCR4 C-terminus contains a fold related to the Mg<sup>2+</sup>-dependent endonuclease core. yCCR4 was also shown to be a component of the major yeast cytoplasmic mRNA deadenylase complex, and to be required for efficient poly(A)-specific mRNA degradation [14,15]. Interestingly, this yCCR4-dependent activity copurifies with yPOP2. Since the presently characterized mammalian yCCR4 ortholog has conserved the ability to bind yPOP2 homologs and displays high sequence similarities with yCCR4 in the nuclease core domain, it suggests that the mammalian yCCR4 orthologs might also be associated with a deadenylase function. The three other characterized families of yCCR4-related proteins have also conserved catalytic residues and structural elements of the Mg<sup>2+</sup>-dependent nuclease core. In addition, their nuclease regions possess residues which are strictly invariant among all the yCCR4-related proteins, from plants to humans, and which are not conserved in any other Mg<sup>2+</sup>-dependent endonuclease-containing protein, thus strengthening the notion of a common and specific function for the yCCR4-related proteins. Hence, according to the present data, proteins of the angel, 3635 or nocturnin families may deserve the same putative function as yCCR4, i.e. deadenylation, or at least share common substrate binding and/or catalytic specificities (for instance preference for RNA and/or poly (A) tail). Along this line, CCR4-specific conserved residues may be important for the setting of such a specificity, whereas the presence -or not- of a LRR domain may determine different associations for these proteins, possibly responsible for differences in cellular localization or in enzymatic activity. Until now, the only known deadenylase in vertebrates is a poly(A)-specific exoribonuclease, PARN, a member of the RNaseD family of exonucleases [34]. The present results therefore propose a series of new putative mammalian deadenylases, and predict a set of residues that should be functionally important for all yCCR4-related proteins.

Finally, the high level of similarity among the C-terminal regions of all paralogous yCCR4-related proteins, along with the divergence of their N-terminal domains, is intriguing in terms of phylogenetic evolution. The C-terminus strong conservation leads to the assumption that all the related genes derive by duplication events from a common ancestor, but the identity of this ancestral gene and the date of expansion can only be hypothesized. Occurrence within the CCR4 family of LRR-containing genes both in yeast and animals, suggests that a LRR-containing gene existed before the divergence of yeast and animals. Whether such a LRR-containing gene could be the common ancestor for all the yCCR4-related genes (with deletion of the LRR and several duplication events

for the angel, 3635 and nocturnin families) is not known. However, the fact that the two *arabidopsis* proteins which cluster with the LRR-containing proteins in the CCR4 branch (Fig. 1A) do not possess a LRR domain, is not in favour of the existence of a LRR-containing ancestor: in that case, at least two independent LRR deletion events (versus one LRR acquisition event, see below) would be required to generate the angel, 3635 and nocturnin families on the one hand, and the CCR4 *arabidopsis* proteins on the other hand. Along this line, BLAST searches in bacteria databases disclosed, in one incompletely sequenced bacterial genome (*Paenibacillus azotofixans*), an open reading frame (nt 3822 to 4583 in the AJ299453 genomic clone) which contains almost all and displays significant similarity with the yCCR4 C-terminal region, but does not contain a LRR nor a N-terminal domain. It is therefore plausible that a "C-terminal core domain" existed first, which was subsequently duplicated several times giving rise to the four identified protein families, and which has been well conserved because it deserved a common enzymatic function, at the opposite of the N-terminal divergent regions. A LRR domain would have been acquired in one of the duplicates, prior to the divergence of animals and fungi.

## Conclusions

The present characterization, in higher eukaryotes, of proteins related to the yeast yCCR4 factor leads to two main conclusions. Firstly, we identified yCCR4 mammalian orthologs and demonstrated that the human member can interact with the human yPOP2 homologs. This result, along with the identification by others of human homologs of NOT proteins, now provides strong evidence for the existence of a human CCR4-NOT complex, with conserved protein-protein interactions. Secondly, we showed that all members of the four yCCR4-related protein families that we have characterized contain the canonical catalytic residues and motifs of the Mg<sup>2+</sup>-dependent endonuclease core, as well as strictly conserved yCCR4-specific residues. These proteins might constitute a new family of deadenylases in mammals, for which critical residues can be predicted.

## Materials and Methods

### Sequence analyses

DNA and amino-acid sequences were examined for homology with the non-redundant nucleotide, protein, and dbEST databases at the NCBI, using the BLAST 2 program. Multiple alignments of amino-acid sequences were carried out using CLUSTALW computer algorithms [23]. The aligned sequences were further analysed, either by the distances method (Neighbor-Joining program) or by the parsimony method (Protpars program, not shown). In the former case, trees were constructed using different parameters, i.e. correction -or not- for

multiple substitutions, and exclusion -or not- of the positions with gaps. The trees obtained under these various conditions were closely related, disclosing only limited variations in the branching pattern of the identified phylogenetic families.

### Characterization of yCCR4-related cDNAs

Characterization of d.nocturnin and dCCR4 cDNAs was achieved by PCR, using two *drosophila* cDNA libraries (an oligo (dT)-primed cDNA library constructed from 0–24 hour-old embryos [gift from C. S. Thummel] and a 3–12 hour-old embryo cDNA library [gift from T. Maniatis]). Characterization of ceCCR4 cDNA was achieved by PCR using a *C. elegans* cDNA library (gift from S. Holbert). The cDNA encompassing the entire mCCR4 open reading frame was cloned by reverse transcription (RT)-PCR reaction, using total RNAs extracted from CBA mouse brain. For RT reactions, one microgram of total RNAs was reverse transcribed from random (dN)6 primers in 20 µl of reaction medium containing 1 mM each deoxynucleoside triphosphate, (dN)6 (2 mM), 20 U of RNAsin, 50 mM Tris-HCl, 75 mM KCl, 5 mM MgCl<sub>2</sub>, and 50 U of Moloney murine leukemia virus reverse transcriptase (Applied Biosystem). Reactions were carried out for 45 min at 42°C. PCR reactions were performed using *rTth* DNA Polymerase (Perkin Elmer), with a 95°C denaturation step for 4 min, and 35 cycles at 58°C for 1 min, 72°C for 3 min, and 94°C for 1 min, and final extension at 72°C for 10 min. Amplification products were subcloned into the T-vector (Promega) and sequenced.

Primers used (indicated in Fig. 1C):

d.nocturnin:5'-CCGATGGATATTGGAAGCTGGG-3',  
 5'-GCAACGGTTGTGTATGAGGCT-3',  
 5'-TCACCACCTCACATATCCCAC-3',  
 5'-TCAGGTACTTGCGGTGCTCCCA-3',  
 5'-GGACTCCCAGGACGATGGCCT-3';  
 dCCR4:5'-GTGCTCTGCGACAAGTACGCGA-3',  
 5'-CGTATAGTTGGTGTGCGGCAT-3',  
 5'-CGCGTACTTGTGCGAGAGCACA-3',  
 5'-CGTTGGCATGCTGACCAGCCT-3',  
 5'-GCGAGTTTTGTCTCCTAAC-3';  
 ceCCR4:5'-GCGGTCGCGACGACAAGAGGA-3',  
 5'-GGAGATCAACACGGTGGACAG-3';mCCR4:  
 5'-ATCTGAGGTCCTCTGAAAGTG-3',  
 5'-GAAGGCGGCGCAGCTCGAGA-3'.

### Mammalian Expression Vectors

The mammalian expression vectors for the GAL4 and VP16 fusions are derivatives of the SV40 promoter-driven expression vector pSG5 (Stratagene). The GAL4 fusion plasmids were obtained by subcloning the appropriate cDNAs into the pGal4PolyII plasmid [24], in-frame with the yeast GAL4 binding domain coding se-

quence. To generate pGal4-hCCR4, a fragment corresponding to the entire hCCR4 coding sequence (with the exception of the initiation codon) was obtained upon digestion of the KIAA1194 cDNA (accession number AB033020, gift from the Kazusa Research Institute, putative start codon -Met 19-positioned by homology with mCCR4) by *SphI* and *MscI* (nt 423 to 2172 in KIAA1194), and was inserted in-frame into the pGal4PolyII vector opened by *XhoI*, after Klenow treatment of both vector and insert. Construction of the deleted fusion protein in which four out of the five hCCR4 leucine-rich repeats are removed (named pGal4-hCCR4( $\Delta$ LRR)) was performed in three steps. The KIAA1194 cDNA was first 3'-truncated upon restriction with *BsmI* (nt 3642) and *SacII* (in pBluescript II SK+ vector), Klenow treatment and self-ligation, to remove undesirable *EcoNI* restriction sites. Deletion of the LRR domain from position 547 to 838 was then achieved upon digestion by *HindIII* (position 547 and 638) and *EcoNI* (position 817 and 838), followed by Klenow treatment and self-ligation. Finally, the deleted open reading frame-containing fragment was excised using *SphI* and *MscI*, and inserted into the pGal4PolyII vector opened by *XhoI*, after Klenow treatment of both vector and insert. To generate pGal4-m.nocturnin, a fragment from the *m.nocturnin* cDNA [21] corresponding to the entire coding sequence (with 9 amino acids upstream of the initiation codon) was excised from pGEX-3X upon *SmaI* and *DraI* digestion, and was inserted into the pGal4PolyII vector opened by *BamHI*, after Klenow treatment of both vector and insert. Expression of the recombinant proteins was checked by in vitro transcription-translation assays (Promega). The pVP16-hCAF1 and pVP16-hPOP2 constructs were obtained upon insertion of the hCAF1 and hPOP2 coding regions into the pSG5FNV vector, in-frame with the VP16 transactivation domain coding sequence, as previously described [17,25].

#### Transfections and luciferase assays

The plasmids used for transfection were prepared by the alkaline/PEG/LiCl method. HeLa cells were grown in DMEM (Gibco/BRL) supplemented with 10% fetal calf serum, seeded at 10 cells/well in 96-well microtiter plates and transfected 8 hours later using Exgen 500 (Euromedex, Souffelweyersheim, France). The DNA for transfection included 100 ng of pG4-TK-Luc reporter plasmid together with 50 ng of GAL4 and/or VP16 fusion vectors, and 10 ng of pCMV-RL vector (cytomegalovirus promoter fused to the renilla luciferase reporter gene [Promega]) as an internal control for transfection efficiency. The amount of SV40 promoter DNA was kept constant upon addition, when necessary, of pSG5 vector to the transfection mixture. The cells were washed and collected 48 h after transfection. Luciferase activity was measured in cell lysates using the Dual Luciferase Kit

(Promega) following the manufacturer's instructions. In all experiments, luciferase activities were normalized with the renilla luciferase activity expressed from the pCMV-RL vector. Each set of experiments was performed in quadruplicate and was repeated at least three times.

#### Far-Western analyses

For in vitro protein-protein interaction assays, 5  $\mu$ g of either GST-CAF1 or GST purified proteins (prepared as described in [25]) were subjected to 10% SDS/PAGE and transferred onto a polyvinylidene difluoride membrane (Millipore) by electroblotting. After denaturation in 6 M and renaturation in 0.187 M guanidine-HCl in HB buffer (25 mM Hepes pH 7.2, 5 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT), the blots were saturated at 4°C in buffer H (20 mM Hepes pH 7.7, 7.75 mM KCl, 0.1 mM EDTA, 25 mM MgCl<sub>2</sub>, 1 mM DTT, 0.05% NP40, 1% dry milk) for 2 hours, and then incubated for 2 hours at 4°C with 50  $\mu$ l of <sup>35</sup>S-methionine-labeled in vitro-translated hCCR4 or luciferase protein (synthesized using a reticulocyte lysate-coupled transcription/translation kit [Promega]). After washing in buffer H for 1 h at 4°C, filters were dried and autoradiographed.

#### Additional material

##### Additional file 1

: *in silico* determined amino acid sequence of d3635protein

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-2-9-S1.doc>]

##### Additional file 2

: *in silico* determined amino acid sequence of h3635 protein

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-2-9-S2.doc>]

##### Additional file 3

: *in silico* determined amino acid sequence of m. angel protein

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-2-9-S3.doc>]

##### Additional file 4

: *in silico* determined amino acid sequence of h. angel protein

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-2-9-S4.doc>]

##### Additional file 5

: *in silico* determined amino acid sequence of ce.angel protein

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-2-9-S5.doc>]

## Acknowledgments

We specially acknowledge S. Nabirochkin and N. Modjtahedi (IGR, Villejuif, France) for the gift of *Drosophila* cDNA libraries, S. Holbert (CEPH, Paris, France) for the gift of a *Caenorhabditis* cDNA library, the Kazusa DNA Research Institute (Kisarazu, Japan) for the gift of the KIAA1194 cDNA clone, and L. Benit and P. Dessen (IGR, Villejuif, France) for their help in the phylogenetic analyses and for discussions. This work was financed by the CNRS and by grants from the ARC.

## References

1. Denis CL: **Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II.** *Genetics* 1984, **108**:833-843
2. Liu H-Y, Toyn JH, Chiang Y-C, Draper MP, Johnson LH, Denis CL: **DBF2, a cell cycle-regulated protein kinase, is physically and functionally associated with the CCR4 transcriptional regulatory complex.** *EMBO J* 1997, **16**:5289-5298
3. Schild D: **Suppression of a new allele of the yeast RAD52 gene by over expression of RAD51, mutations in src2 and ccr4, or mating-type heterozygosity.** *Genetics* 1995, **140**:115-127
4. McKenzie EA, Kent NA, Dowell SJ, Moreno F, Bird LE, Mellor J: **The centromere promoter factor I, CPFI, of *Saccharomyces cerevisiae* modulates gene activity through a family of factors including SPT21, RPD1 (SIN3), RPD3, and CCR4.** *Mol. Gen. Genet* 1993, **240**:374-386
5. Draper MP, Liu H-Y, Nelsbach AH, Mosley SP, Denis CL: **CCR4 is a glucose-regulated transcription factor whose leucine-rich repeat binds several proteins important for placing CCR4 in its proper promoter context.** *Mol. Cell. Biol* 1994, **14**:4522-4531
6. Liu H-Y, Badarinarayana V, Audino DC, Rappsilber J, Mann M, Denis CL: **The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively.** *EMBO J* 1998, **17**:1096-1106
7. Bai Y, Salvatore C, Chiang Y-C, Collart MA, Liu H-Y, Denis CL: **The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5.** *Mol. Cell. Biol* 1999, **19**:6642-6651
8. Liu H-Y, Chiang Y-C, Pan J, Chan J, Salvatore C, Audino DC, Badarinarayana V, Palaniswamy V, Anderson B, Denis CL: **Characterization of CAF4 and CAF16 reveal a functional connection between the CCR4-NOT complex and a subset of SRB proteins of the RNA polymerase II holoenzyme.** *J. Biol. Chem* 2001, **276**:7541-7548
9. Chang M, French-Cornay D, Fan H-Y, Klein H, Denis CL, Jaehning JA: **A complex containing RNA polymerase II, Paf1p, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signaling.** *Mol. Cell. Biol* 1999, **19**:1056-1067
10. Malvar T, Biron RW, Kaback DB, Denis CL: **The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of ADH2 gene expression.** *Genetics* 1992, **132**:951-962
11. Draper MP, Salvatore C, Denis CL: **Identification of a mouse protein whose homolog in *Saccharomyces cerevisiae* is a component of the CCR4 transcriptional regulatory complex.** *Mol. Cell. Biol* 1995, **15**:3487-3495
12. Diakic M: **Functionally unrelated signalling proteins contain a fold similar to Mg<sup>2+</sup>-dependent endonucleases.** *TIBS* 2000, **25**:272-273
13. Hofmann K, Tomiuk S, Wolff G, Stoffel W: **Cloning and characterization of the mammalian brain-specific, Mg<sup>2+</sup>-dependent neutral sphingomyelinase.** *Proc. Natl. Acad. Sci* 2000, **97**:5895-5900
14. Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R: **The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*.** *Cell* 2001, **104**:377-386
15. Daugeron MC, Mauxion F, Séraphin B: **The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation.** *Nucleic Acids Res* 2001, **29**:2448-2455
16. Bogdan JA, Adams-Burton C, Pedicird DL, Sukovich DA, Benfield PA, Corjay MH, Stoltenborg JK, Dicker IB: **Human carbon catabolite repressor protein (CCR4)-associative factor I: cloning, expression and characterization of its interaction with the B-cell translocation protein BTG1.** *Biochem. J* 1998, **336**:471-481
17. Rouault J-P, Prévôt D, Berthet C, Birot A-M, Billaud M, Magaud J-P, Corbo L: **Interaction of BTG1 and p53-regulated BTG2 gene products with mCaf1, the murine homolog of a component of the yeast CCR4 transcriptional regulatory complex.** *J. Biol. Chem* 1998, **273**:22563-22569
18. Fidler C, Wainscoat JS, Boulwood J: **The human POP2 gene: identification, sequencing, and mapping to the critical region of the 5q- syndrome.** *Genomics* 1999, **15**:134-136
19. Albert TK, Lemaire M, van Berkum NL, Gentz R, Coolart MA, Timmers HTM: **Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits.** *Nucleic Acids Res* 2000, **28**:809-817
20. Benson JD, Benson M, Howley PM, Struhl K: **Association of distinct yeast Not2 functional domains with components of Gcn5 histone acetylase and Ccr4 transcriptional regulatory complex.** *EMBO J* 1998, **17**:6714-6722
21. Dupressoir A, Barbot W, Loireau M-P, Heidmann T: **Characterization of a mammalian gene related to the yeast CCR4 general transcription factor and revealed by transposon insertion.** *J. Biol. Chem* 1999, **274**:31068-31075
22. Green CB, Besharse JC: **Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin.** *Proc. Natl. Acad. Sci* 1996, **93**:14884-14888
23. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680
24. Green S, Issemann I, Sheer E: **A versatile in vivo and in vitro eukaryotic expression vector for protein engineering.** *Nucleic Acids Res* 1988, **16**:369
25. Prévôt D, Morel AP, Voeltzel T, Rostan MC, Rimokh R, Magaud JP, Corbo L: **Relationships of the antiproliferative proteins BTG1 and BTG2 with CAF1, the human homolog of a component of the yeast CCR4 transcriptional complex: involvement in estrogen receptor alpha signaling pathway.** *J. Biol. Chem* 2001, **30**:9640-9648
26. Kurzik-Dumke U, Zengerle A: **Identification of a novel *Drosophila* melanogaster gene, angel, a member of a nested gene cluster at locus 59F4.5.** *Bioch. Biophys. Acta* 1996, **1308**:177-181
27. Nagase T, Ishikawa K, Kikuno R, Hirokawa M, Nomura N, Ohara O: **Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.** *DNA res* 1999, **6**:337-345
28. Buchanan SG, Gay NJ: **Structural and functional diversity in the leucine-rich repeat family of proteins.** *Prog. Biophys. Molec. Biol* 1996, **65**:1-44
29. Suck D, Oefner C: **Structure of DNaseI at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA.** *Nature* 1986, **321**:620-625
30. Mol CD, Kuo CF, Thayer MM, Cunningham RP, Tainer JA: **Structure and function of the multifunctional DNA-repair enzyme exonuclease III.** *Nature* 1995, **374**:381-386
31. Gorman MA, Morera S, Rothwell DG, de la Fortelle E, Mol CD, Tainer JA, Hickson ID, Freemont PS: **The crystal structure of the human DNA repair endonuclease HAPI suggests the recognition of extra-helical deoxyribose at DNA abasic sites.** *EMBO J* 1997, **16**:6548-6558
32. Barzilay G, Walker LJ, Robson CN, Hickson ID: **Site-directed mutagenesis of the human DNA repair enzyme HAPI: identification of residues important for AP endonuclease and RNase H activity.** *Nucleic Acids Res* 1995, **23**:1544-1550
33. Ikematsu N, Yoshida Y, Kawamura-Tsuzuku J, Ohsugi M, Onda M, Hirai M, Fujimoto J, Yamamoto T: **Tob2, a novel anti-proliferative Tob/BTG1 family member, associates with a component of the CCR4 transcriptional regulatory complex capable of binding cyclin-dependent kinases.** *Oncogene* 1999, **18**:7432-7441
34. Korner CG, Wahle E: **Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease.** *J. Biol. Chem* 1997, **272**:10448-10456