LUND UNIVERSITY

The G-protein coupled receptor CMKLR1/ChemR23: Studies on gene regulation, receptor ligand activation, and HIV/SIV co-receptor function

Mårtensson, Ulrika

2005

Link to publication

Citation for published version (APA):
Mårtensson, U. (2005). *The G-protein coupled receptor CMKLR1/ChemR23: Studies on gene regulation, receptor ligand activation, and HIV/SIV co-receptor function*. [Doctoral Thesis (compilation)]. Elsevier.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Gene 328 (2004) 167–176

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Genomic organization and promoter analysis of the gene encoding the mouse chemoattractant-like receptor, CMKLR1

Ulrika E.A. Mårtensson*, Christer Owman, Björn Olde

Division of Molecular Neurobiology, Wallenberg Neuroscience Center, BMC A12, SE-221 84, Lund, Sweden

Received 1 August 2003; received in revised form 6 November 2003; accepted 2 December 2003

Received by S.M. Mirkin

Abstract

Chemoattractant-like receptor 1 (CMKLR1) is a functionally unknown (‘‘orphan’’) G-protein coupled receptor. It has been implicated in osseous and cartilage development, and it also has a pathophysiological role as one of the minor coreceptors involved in human immunodeficiency virus type I (HIV-1)/simian immunodeficiency virus (SIV) infection of $CD4^+$ immune cells. Here we report the cloning of the mouse *cmklr1* gene, the characterization of its genomic structure for comparison with the human gene, and the mapping and functional analysis of its 5′ flanking sequence. The gene was found to contain three exons intercepted by one larger and one smaller intron. The overall structure resembles the human orthologue. The promoter lacks classical TATA and CCAAT boxes but contains several GC-rich regions as well as AP-4 elements, C/EBP motifs, and GATA-1 and GATA-2 binding sites. Promoter function was analyzed in mouse neuroblastoma (NB4 1A3) cells, endogenously expressing CMKLR1, as well as in mouse embryonic fibroblastic (3T3 clone A31) cells not expressing CMKLR1. 5′ Deletion analysis and luciferase reporter gene assays of the promoter indicated that a 280-bp sequence adjacent to the transcription start site (established through 5′-RACE) is essential for initiating transcription. Within this region it was possible to identify four potential Sp1-binding sites that may be active in the transcription of the gene. Thus, we show that the m*cmklr1* gene has several conserved features in common with its human counterpart, which suggests that they are regulated in a similar manner. The promoter does not seem to be tissue specific but other elements or enhancers may be missing. The results provide a necessary basis for further studies of the gene regulation and function of this chemoattractant-like receptor and will be useful when manipulating the gene in the development of transgenic animal models.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Orphan G-protein coupled receptor; Transcriptional regulation; Luciferase reporter gene; Sp1

1. Introduction

G-protein coupled receptors (GPCRs) belong to a superfamily and are characterized by their localization to the cell membrane and general structure of seven membrane-spanning regions. The receptors are activated by a broad spectrum of highly divergent substances (Watson and Arkinstall, 1994). One subfamily of GPCRs comprises the ‘‘classical leukocyte chemoattractant receptors’’, exemplified by the receptors for complement factor, *N*-formyl peptide, and leukotriene B4 (Murphy, 1994). During inflammation these receptors become activated and contribute to the host–defence reaction to microbes and antigens.

Chemoattractant-like receptor 1 (CMKLR1) is a functionally unknown (i.e., orphan) GPCR with high homology to other chemoattractant-type receptors (Gantz et al., 1996). The mouse orthologue, mCMKLR1, has previously been described under the name DEZ by Methner et al. (1997); it was originally cloned from a neuroblastoma/glioma cell line and from a cDNA library of adult mouse brain (Methner et al., 1997). In situ hybridisation was used to show that the receptor is differently regulated during embryonic development, with high expression in cartilage and osseous tissue, whereas the main expression in the adult mouse was seen in parathyroid glands, lungs, and in blood vessels of the choroids plexus (Methner et al., 1997).

Abbreviations: Sp1, stimulating protein 1; AP-4, activator protein 4; C/EBP, ccaat/enhancer binding protein; GATA-1, GATA-binding factor 1; GATA-2, GATA-binding factor 2; cDNA, complementary DNA; PCR, polymerase chain reaction; TBE, Tris borate/EDTA; bp, base pair(s); kb, kilobase(s); ORF, open reading frame.

* Corresponding author. Tel.: +46-46-2220521; fax: +46-46-2220568.
E-mail address: Ulrika.Martensson@mphy.lu.se (U.E.A. Mårtensson).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2003.12.004

The rat and the human orthologues, rCMKLR1 and hCMKLR1, have also been described under the name CMKRL3 (Owman et al., 1997) and ChemR23 (Samson et al., 1998), respectively. Interestingly, hCMKLR1 was found to be expressed in dendritic cells and macrophages. It turned out to function as a coreceptor for select isolates of simian immunodeficiency virus (SIV) and human immunodeficiency virus-1 (HIV-1) facilitating their entry into $CD4^+$ cells, thereby assisting the major viral coreceptors, CCR5 and CXCR4 (Samson et al., 1998). As part of a project aimed at elucidating the physiological and pathophysiological roles of CMKLR1 we have isolated the m*cmklr1* gene, mapped its genomic structure, and characterized its promoter in order to clarify the molecular mechanisms behind the gene regulation of mCMKLR1.

2. Materials and methods

2.1. Cell lines

The mouse neuroblastoma cell line, NB4 1A3 (ECACC, no. 89121405), and the mouse embryonic fibroblastic cell line, 3T3 clone A31 (ECACC, no. 86110401), were maintained in Dulbecco's modified Eagle's medium + Glutamax I (Invitrogen) supplemented with 10% fetal bovine serum (FBS) and 0.5% penicillin + streptomycin. In addition, the NB4 1A3 medium contained 1% sodium pyruvate (Sigma). The cells were incubated at 37 °C in 7% CO_2.

2.2. Northern blot analysis

Northern blot analysis of a premade mouse cell line blot (2 μg mRNA/lane) was performed according to the manufacturer's (Clontech) instructions. Total RNA was isolated from the 3T3 cells by the guanidinium isothiocyanate method (Chomczynski and Sacchi, 1987), mRNA was selected using a commercial kit (Amersham Biosciences), and 6 μg was used for the blot. A probe containing the coding region of m*cmklr1* was radioactively labelled with [α-P^{32}]dCTP (NEN) using the Megaprime DNA labelling Kit (Amersham Biosciences). The blot was hybridized and washed according to standard procedures (Sambrook et al., 1989) and exposed to X-ray film (Kodak) overnight at − 70 °C.

2.3. Rapid amplification of cDNA ends (5′-RACE)

Total RNA and mRNA were isolated from NB4 1A3 cells as described in section 2.2. 5′-RACE was performed using the Marathon™ cDNA Amplification kit (Clontech). In the primary and secondary PCR the adaptor-specific primers (provided in the kit) were used together with the m*cmklr1*-specific primers, 5′GACCCAGACGGCCGAGCAGGTCATGTAG 3′, and 5′GACCAGACGGGGAGCAGCACGGAGATGC 3′, respectively. The amplified 5′-fragment was subcloned into pBluescript (SK-) for sequence analysis using the BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems).

2.4. Mapping of the mcmklr1 gene

To isolate the genomic region containing the m*cmklr1* gene, a mouse bacterial artificial chromosome (BAC) library (mouse strain 129/SvJ) was screened by Genome Systems using a probe corresponding to the coding region of the r*cmklr1* (Owman et al., 1997). A positive, approx. 200-kb BAC clone was purified using the Nucleobond PC 500 kit (Macherey-Nagel) according to the manufacturer's description. Isolation of the m*cmklr1* gene was performed by restriction digestions, southern blotting, and by PCR-based genome walking using the Universal GenomeWalker™ kit (Clontech). Positive fragments were subcloned into pBluescript (SK-) (Stratagene) for restriction mapping and sequencing as above.

2.5. Construction of reporter plasmids for analysis of promoter activity

The region immediately upstream the 5′ end of the m*cmklr1* transcription start was isolated by genome walking (Universal GenomeWalker™ kit, Clontech) using BAC DNA as template. In the primary and secondary PCR the adaptor-specific primers (provided in the kit) were used together with the m*cmklr1* primers, 5′ CTCAAACCCTGATTTCGCAGGAGCCGG 3′ and 5′ CAGAACAACTGAGAAAGAGGCCAGAGC 3′, respectively. The amplified products were subcloned into the luciferase reporter plasmid, pGL3-Enhancer (Promega).

2.6. 5′ Deletions of the promoter region

5′ Deletions of the promoter construct were performed using the Erase-A-Base kit (Promega). The promoter plasmid was linearized with *Mlu*I and *Kpn*I, and the fragment was purified on a column (Qiagen). The linearized plasmid was digested with exonuclease III at 30 °C for different lengths of time (Henikoff, 1984). The deletions were confirmed by sequence analysis.

2.7. Luciferase reporter assay

The day before transfection, 6×10^4 NB4 1A3 cells/well were seeded in white 96-well tissue culture plates (Costar). Luciferase constructs (90 ng) were co-transfected with pRL-TK (3 ng) (Promega) as internal control, using 0.5 μl of the transfection agent, TransIT-LT1 (Mirus). Forty-two hours after transfection, the cells were harvested in 20-μl reporter lysis buffer (Promega) and stored at − 70 °C until analysis. Firefly and *Renilla* luciferase activities were measured in a dual-luciferase assay (Promega) using a BMG Lumistar microplate luminometer. Experiments were performed at least three times in hextuplicates.

Table 1
Oligonucleotides used as probes in gel-shift and super-shift assays[a]

Name	Sequence in 5′ to 3′ direction (sense strand)[b]	Location[c]
Sp1 cons	ATTCGATC**GGGGCGGGGC**GAGC	
1Sp1 wt	TGCGCCCAGACAA**GGGGCGGG**TCATTCCTAGTGGG	– 205/ – 171
1Sp1 mut	TGCGCCCAGACAAGGttaGGGTCATTCCTAGTGGG	
2Sp1 wt	TCTGGGGGAGT**GGCGGGGGGC**CCACTCCACACAGC	– 125/ – 91
2Sp1mut	TCTGGGGGAGTttaGGGGGGCCCACTCCACACAGC	
3Sp1wt	CCCACTCCACACAGCT**GGCGCCCCCGCCC**TCCCCAGGCTGTTTC	– 105/ – 62
3Sp1mut	CCCACTCCACACAGCTGGCGCCCCttaCCTCCCCAGGCTGTTTC	
4Sp1wt	TCCTGCTCCTCCC**GCCCCTCCCCGC**TCCAGCTGTTTCTCAGGCT	– 63/ – 20
4Sp1mut	TCCTGCTCCTCCCGCttaTCCCCGCTCCAGCTGTTTCTCAGGCT	

[a] All oligonucleotides were synthesized by MWG Biotech.
[b] Sp1 binding motifs are indicated in bold letters. Mutations relative to the corresponding wild-type (wt) sequence are shown in lowercase letters.
[c] Locations of oligonucleotides are presented as base pairs upstream of the transcription start site.

2.8. Electrophoretic mobility shift assay (EMSA)

Nuclear extracts were prepared from NB4 1A3 cells essentially as previously described (Andrews and Faller, 1991), with the following modifications: 0.5 mM DTT (dithiothreitol) + 0.2 mM PMSF (phenylmethylsulphonyl fluoride) in buffer A and C were replaced by protease inhibitor tablets (Roche), and buffer A was supplemented with 10% Nonidet P40 (Roche). Protein concentrations were determined using the BCA Protein Assay Reagent kit (Pierce). Synthetic oligonucleotides (Table 1) were used as probes in the gel-shift and super-shift assays. Sense strands were labelled with [γ-P^{32}]ATP (Amersham Biosciences) using the 5′ end labelling kit (Amersham Pharmacia Biotech) and purified on Microspin™ G-25 columns (Amersham Pharmacia Biotech). The labelled sense strands were mixed with 2 × molar excess of antisense strands in 10 mM Tris–HCl (pH 7.6). After denaturation by boiling for 2 min in a water bath, the strands were allowed to anneal by cooling the bath to room temperature. Binding was performed in a 15-μl reaction mixture containing 10 fmol P^{32} labelled probe, 10 mM Tris–HCl (pH 7.5), 7 mM $MgCl_2$, 8% glycerol, 1 μg poly-(dI–dC) (Amersham Biosciences), and 6 μg of nuclear extract. The mixture was incubated at 25 °C for 30 min. Unlabelled DNA competitors were mixed with nuclear extracts and pre-incubated at 25 °C for 20 min before labelled probe was added and incubated at 25 °C for another 30 min. The DNA–protein complexes formed were resolved on a 6% nondenaturing polyacrylamide gel. Electrophoresis was carried out in 1 × TBE buffer at 150 V for 3 h at 4 °C. The dried gel was exposed to X-ray film (Kodak) overnight at – 70 °C. For super-shift assays, 4 μg of a polyclonal Sp1 antibody (Santa Cruz Biotechnology, cat. nr. sc-59X) and extracts were incubated overnight at 4 °C whereafter probe was added and the incubation extended for an additional 30 min at 25 °C.

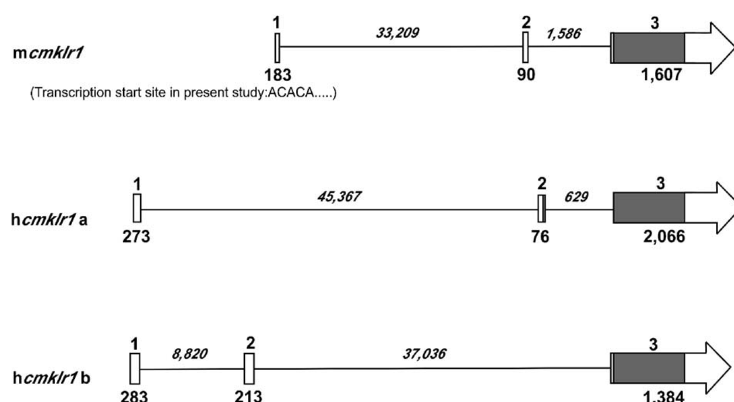

Fig. 1. Schematic genomic organization of the m*cmklr1* gene, and comparison to the two variants of h*cmklr1* found in GenBank (accession no. U79526 for h*cmklr1*a and U79527 for h*cmklr1*b). The exons (1–3, lengths in roman figures) are presented as boxes. Shaded boxes show the coding region while open boxes indicate untranslated regions. The mouse gene is located to chromosome 5, region F, while the human gene is located to chromosome 12, regions q21.1–24.1. Intron lengths are shown in italic figures. Sizes of exons and introns are given in number of base pairs.

Table 2
Putative exon–intron organization of mouse and human *cmklr1*[a]

Exon number	Species	Size (bp)	5′-Donor site	Intron size (bp)	3′-Acceptor site
1	m*cmklr1*	183	GAGAAG gtacag	33,209	ctgcag GTATTT
	h*cmklr1*a	273	GGGTGT gtatca	45,367	ctacag GTGTTT
	h*cmklr1*b	283	GGGAAG gtatca	8,820	ctgtag CCCATC
2	m*cmklr1*	90	CCAAAG gtaggg	1,586	tcacag AGATGG
	h*cmklr1*a	76	TGAATG gtgagg	629	ttgcag AGAATG
	h*cmklr1*b	213	TGACAT gtgagt	37,036	ttgcag AGAATG
3	m*cmklr1*	1607			
	h*cmklr1*a	2066			
	h*cmklr1*b	1384			

[a] Exon sequences are shown in uppercase letters while intron sequences are shown in lowercase letters. Underlined sequences represent consensus nucleotides typical for exon–intron boundaries.

2.9. Data analysis

Luciferase reporter experiments were performed as hextuplicates where each experiment was repeated three times. Statistical errors are shown as standard deviations (SD). Unpaired Student's *t*-test was performed to determine statistical significance using GraphPad Prism (GraphPad Software, San Diego, CA). Sequence alignments were carried out using the ClustalX program.

3. Results

3.1. Genomic organization of mouse cmklr1 gene and comparison to human cmklr1

The genomic region containing m*cmklr1* was isolated by screening of a mouse BAC library using a probe corresponding to the coding region of the rat orthologue, r*cmklr1*. Alignment of m*cmklr1* cDNA from NB4 1A3

```
mCMKLR1     1  --MEYDAYNDSGIYDDEYSDGFGYFVDLEEASPWEAKVAPVFLVVIYSLVCFLGLLGNGL
rCMKLR1     1  --MEYEGYNDSSIYGEEYSDGSDYIVDLEEAGPLEAKVAEVFLVVIYSLVCFLGILGNGL
hCMKLR1a    1  MRMEDEDYNTSISYGDEYPDYLDSIVVLEDLSPLEARVTRIFLVVVYSIVCFLGILGNGL
hCMKLR1b    1  --MEDEDYNTSISYGDEYPDYLDSIVVLEDLSPLEARVTRIFLVVVYSIVCFLGILGNGL

mCMKLR1    59  VIVIATFKMKKTVNTVWFVNLAVADFLFNIFLPMHITYAAMDYHWVFGKAMCKISNFLLS
rCMKLR1    59  VIVIATFKMKKTVNTVWFVNLAVADFLFNIFLPIHITYAAMDYHWVFGKAMCKISSFLLS
hCMKLR1a   61  VIIIATFKMKKTVNMVWFLNLAVADFLFNVFLPIHITYAAMDYHWVFGTAMCKISNFLLI
hCMKLR1b   59  VIIIATFKMKKTVNMVWFLNLAVADFLFNVFLPIHITYAAMDYHWVFGTAMCKISNFLLI

mCMKLR1   119  HNMYTSVFLLTVISFDRCISVLLPVWSQNHRSIRLAYMTCSAVWVLAFFLSSPSLVFRDT
rCMKLR1   119  HNMYTSVFLLTVISFDRCISVLLPVWSQNHRSVRLAYMTCVVVWVLAFFLSSPSLVFRDT
hCMKLR1a  121  HNMFTSVFLLTIISSDRCISVLLPVWSQNHRSVRLAYMACMVIWVLAFFLSSPSLVFRDT
hCMKLR1b  119  HNMFTSVFLLTIISSDRCISVLLPVWSQNHRSVRLAYMACMVIWVLAFFLSSPSLVFRDT

mCMKLR1   179  AN-IHGKITCFNNFSLAAPESSPHPAHSQVVSTGYSRHVAVTVTRFLCGFLIPVFIITAC
rCMKLR1   179  VSTSHGKITCFNNFSLAAPEPFSHSTHPRTDPVGYSRHVAVTVTRFLCGFLIPVFIITAC
hCMKLR1a  181  AN-LHGKISCFNNFSLSTPGSSSWPTHSQMDPVGYSRHMVVTVTRFLCGFLVPVLIITAC
hCMKLR1b  179  AN-LHGKISCFNNFSLSTPGSSSWPTHSQMDPVGYSRHMVVTVTRFLCGFLVPVLIITAC

mCMKLR1   238  YLTIVFKLQRNRLAKNKKPFKIIITIIITFFLCWCPYHTLYLLELHHTAVPSSVFSLGLP
rCMKLR1   239  YLTIVFKLQRNRLAKTKKPFKIIITIIITFFLCWCPYHTLYLLELHHTAVPASVFSLGLP
hCMKLR1a  240  YLTIVCKLHRNRLAKTKKPFKIIVTIIITFFLCWCPYHTLNLLELHHTAMPGSVFSLGLP
hCMKLR1b  238  YLTIVCKLQRNRLAKTKKPFKIIVTIIITFFLCWCPYHTLNLLELHHTAMPGSVFSLGLP

mCMKLR1   298  LATAVAIANSCMNPILYVFMGHDFRKFKVALFSRLANALSEDTGPSSYPSHRSFTKMSSL
rCMKLR1   299  LATAVAIANSCMNPILYVFMGHDFKKFKVALFSRLVNALSEDTGPSSYPSHRSFTKMSSL
hCMKLR1a  300  LATALAIANSCMNPILYVFMGQDFKKFKVALFSRLVNALSEDTGHSSYPSHRSFTKMSSM
hCMKLR1b  298  LATALAIANSCMNPILYVFMGQDFKKFKVALFSRLVNALSEDTGHSSYPSHRSFTKMSSM

mCMKLR1   358  NEKASVNEKETSTL
rCMKLR1   359  IEKASVNEKETSTL
hCMKLR1a  360  NERTSMNERETGML
hCMKLR1b  358  NERTSMNERETGML
```

Fig. 2. Amino acid sequence alignment of the CMKLR1 orthologues in mouse (accession no. U79525), rat (accession no. AJ002745), and human (accession no. U79526 for hCMKLR1a and U79527 for hCMKLR1b) was performed using the ClustalX program. Shaded boxes represent regions of identity between the species. The seven putative TM regions (I–VII) in mouse were assessed using the program TMHMM 2.0 and indicated by horizontal lines.

Fig. 3. Northern blot analysis of a mouse cell line blot, where each well was loaded with 2 μg mRNA. The size of the major transcript is indicated with arrow. The mouse cell line used as a negative control in luciferase assays (3T3 clone A31) is shown in a separate lane, loaded with 6 μg mRNA. Signals from the β-actin probe are shown below.

cells with genomic sequences obtained from the BAC clone revealed the exon–intron organization shown in Fig. 1 (sequence data from this study have been deposited in GenBank under accession nos. AY342407, AY 342408 and AY342409). The size of the large intron 1 was estimated by restriction cleaving of cloned fragments and PCR-based genome walking. The genomic structure of m*cmklr1* could be confirmed when the sequence of the m*cmklr1* locus, localized to region F of mouse chromosome 5, was in the course of this study presented in the mouse genome database. Alignment of genomic sequences and the m*cmklr1* mRNA deposited in GenBank by Methner et al. (1997) (accession no. U79525) showed a slightly different organization in exon 1 in that the reported transcription start site is differently located (Fig. 4B).

For comparison with the mouse genomic organization, the organization of the human orthologue, h*cmklr1*, was obtained by alignment of genomic sequences from the human genome database (unfinished contig. NT 035235.2: 4521738–4570155) and the three mRNA transcripts deposited in GenBank (accession nos. U79526, U79527 and NM004072). The h*cmklr1* locus is localized to human chromosome 12, region q.21.2–24.1 (Gantz et al., 1996; Samson et al., 1998). The comparison of the genomic organization is schematically illustrated in Fig. 1. Both genes span over extensive genomic regions containing large introns. The m*cmklr1* gene spans 36,675 bp and consists of three exons, where the first and the second contain only untranslated sequence, while the coding region is localized to the third exon. The two introns are 33,209- and 1586-bp long. The exon/intron boundaries of the mouse and the human genes are presented in Table 2.

3.2. Species homology

Protein sequence alignment of CMKLR1 orthologues in mouse (accession no. U79525), rat (accession no. AJ002745) and human (accession no. U79526 for hCMKLR1a and U79527 for hCMKLR1b) showed a strong similarity between species (Fig. 2). The mouse and the two human splicing variants were 70% identical, the human splicing variants a and b showing 99% amino acid sequence identity. The mouse and the rat orthologues (the latter located to chromosome 12, region 12q16) displayed 91% identity.

Fig. 4. (A) Identification of the transcription start site of m*cmklr1* by 5′-RACE using cDNA from NB4 1A3 cells as template. This revealed a 650-bp fragment that was subcloned and sequenced. The size marker lane with two relevant band sizes is indicated. (B) Sequencing of six different subclones revealed four possible transcriptional start sites. The start site published by Methner et al. (1997) is shown for comparison.

A

```
-1039 TAAGTAGCTT TAGGCAGCAG GGCTTCCAAC CATGACAGAC CTCAGCTTGG TCTTAGCTGT
                AP-4                              AP-4
-979  GAGCCCGAGC CGGCTGCCTG GGCACAGATT CTGGCCCACG GCCTCAGCAT CTCCATCCGC
                                                     GATA-1   GATA-2
-919  CATGCGTGAT CTACTCTTCC AGCCCCGTTG CAGCAGGCAT GGTGATGCAT GACACTCTCA
                                                 GATA-1
-859  ACCAAAGCAC TTGCTATGGA TCACACAGGC CAAGGACTGT TATGTCTTCC TTCTGGAGGT
-799  TGAGGACTGG GATCTAGCCT GGTTGGGCTC CAGCTCTGCC ATCTCCTAGC TGGTAGCAT
                                        GATA-1/GATA-2
-739  AGGGTAAGAA ACTTAACCTT TCTGTGCCCC AATTTCCTTA CCAATAAACA AGGGTAATAA
                                           C/EBP
-679  AAATACCTGT CCACAACCAC CACTGCCGTA GGGAGGATGT AGTATCAAGT TCCTAGCACT
-619  GTGAAGGTTG TGTACTGAGT GCCCAGTACC CTGCCTTGTT CTTGGCTACT TCAACAATTC
                                Sp1
-559  TGTTGCTGTC GGTGCCAATA AACTAGGCAC TATATCAGGT GGTGGGGACA GACTGGGGAA
                                     GATA-1
-499  GACAGGCAGC AGGCAGGACA GCGTGATACC TCTGTCAGTG CTTGCTGTGA TGTGGAAACC
                           GATA-1/GATA-2                GATA-1  C/EBP
-439  CAGAGTAGTC TGTCAGGGGA GGTTTCTTCA TCACGGAGAA GGAGCTGGAA GCCATCTGGC
                                GATA-1                    GATA-1/GATA-2
-379  CACACTAGGG GCAAGAATAG CACTCAGAAG TAGGGAACAG CAGGCAAGAG GCCATAGGGC
-319  TTGCTACCGT CCCTACTGCG GGCAAAGATA ACCAGCTGCG GGTGGGTCTT CAGGTCGGTC
                           GATA-1/GATA-2  AP-4
-259  AGGAGTAACT TCCCCAGCAG CCTGGTTTCA AGGCTCCATT TGCGGCTCCA GTTGTGCGCC
                                                                  1Sp1
-199  CAGACAAGGG GCGGGTCATT CCTAGTGGGC CGTTACCACC CAGGAGCCCA GCCTGGCTGG
              Sp1                        C/EBP
-139  CAGGCGAAGC CTCCTCTGGG GGAGTGGCGG GGGGCCCACT CCACACAGCT GGCGCCCCCG
                    2Sp1      Sp1     3Sp1       AP-4       Sp1
-79   CCCTCCCCAG GCTGTTTCCT GCTCCTCCCG CCCCTCCCCG CTCCAGCTGT TTCTCAGGCT
                       4Sp1            Sp1          AP-4
-19   GAGTTCTGTC CCTGTCCCCA CACAGAACAC AGGACAGAGG GAGGCTCTTA GGATGTTGTG
                          +1
41    CTCCGCGGGG CTCAGACGAA ATCTTCTGTG AATGGAAGAA ATGCTTCCAA GCAAACAGCC
101   ACTACCAGAA TGAGAAAGAG GCCAGAGCGC GAGTTCTCAA ACCCTGATTT CGCAGGAGCC
161   GGAGGGGGAT ATTGGAGAGA AGGTATTTCC AGTCACGCGC AGTAACAGAC CAGCCAAGGA
221   CCAGGACTGG AGTTCTGTTC TACAACGGTG GAACAGTGAA CGGTCTCCAA AGAGATG
                                                              +275
```

B

```
-1017 ATCCTCATGT GTAGAATGGG CAAGTGATGC CCCCCCGACG TGGGTGGTGG TTCACACGGC
-967  TTACTCCATG GAAGCACTTG GCACAGGGGC CCACATGGGC ACTTCTGTTA TTTCTTCAGT
-907  CTGGAGGCCA AGGGTGGGCC TCCGAGCCAG ACAACATAGT TGGAATCCCA GCTCTGTCAG
                                                                AP-4
-847  CTGCTGGCGG TGTGGTCTTG AGCAAGTGAC TTAATCTCTC TGTACCCCAG TTTCCTTATT
                                                                C/EBP
-787  GGTAAGACTG GAGTAGTAAT AATTCCTGTT ATCTAGGGTT GCCGTGAAGA GTAAGTACGC
                                 GATA-1/GATA-2
-727  ATCGCCTAGC ACGCTGCCTG GCTGTCGGTT GTTTTGATTA TTCCACTAAT GATGGTGACC
                 GATA-1                              GATA-1   C/EBP
-667  ACACAGCCGA TGAACCAGGC ACTTAGCGCC TGGTTAGGGC AGGGAACAGA ATACAGTTCT
-607  TTCCCTACAG AGTTCCTGGT GCAAGGAGTT CTCGGTGCAA GGAGAGAGAC ATCTTAACAG
-547  GCAAGGACAG CACTGCATCA GTGAGGTGAT GGGGTGACCC TGAGGGGCAC AGCAGCCCCC
                               GATA-1/GATA-2
-487  GAGGGTTGGG GAAGGCTTCC CTGGGAGGTG ACCTTCAACC TGAGACCTGA AGAGGCAGTA
-427  GGAGACCATC GGGCCACGAA TTAGGGCAGG AATAGCATCC AGGTGGGAGG AACAGCGTTG
         GATA-1/GATA-2          Sp1
-367  CAAAGGCCCA GGGTGAGAGG CAGCCAATCT CTCGGTCAGT GTGGAGCTTG AGCCCCTCAC
-307  CGGGAGCAGA GGCGCTTGCA CTTCCCCCAG CAGCCGGGTG CTCAGATGTG CTCCCTGTTT
-247  GCGCTCAGTG GGCACGGCTC CCCGGCTGCA AGAGGCTGCC AGAGGGGCGG GTCGTCCCCA
                                                    Sp1
-187  GCGGGCCCCC GCCGCCACCC ACAAGCCCAG TCCCCGGCTG GTAGGGGCAC CCTCCCCTCG
-127  CAGCCCCCTC TCCACGCAGC TAGCCGCCCC CGCCCTAGGT CAGCCCAGTT TCCTGCCCCT
                                  Sp1
-67   CCCTCCTGCT CTCCTCCCTC CAGCTGTTTC TCTGGCTGAG TTCTGTCCCA GTCCCATTCA
                          AP-4                                     +1
                                                                hcmklr1b
-7    GGGAAGCA
             +1
          hcmklr1a
```

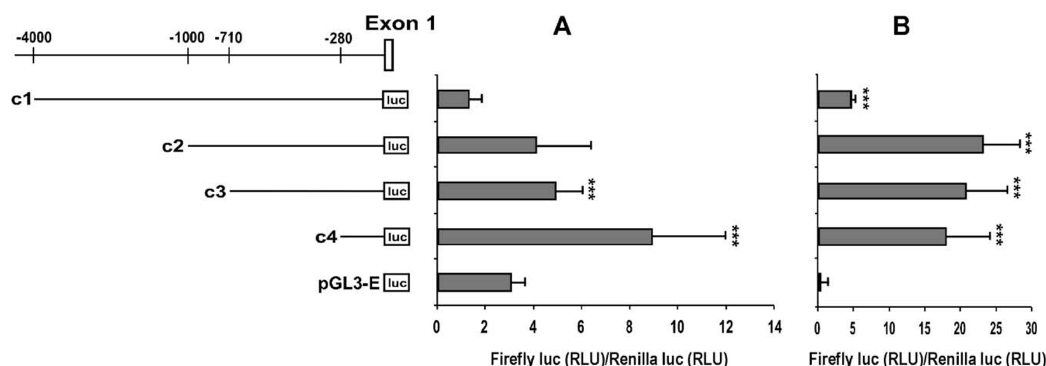

Fig. 6. Deletion analysis of the 5′ flanking region upstream of exon 1 of m*cmklr1* (indicated, fragment sizes being shown in number of base pairs). Deletion constructs (c1–4), generated with exonuclease III digestion, were transiently transfected into NB4 1A3 cells (A) and 3T3 clone A31 cells were used as negative control (B). Forty-two hours after transfection, cells were harvested and assayed for luciferase activity. A comparison between the response with empty plasmid (pGL3-E) and various deletion constructs, *** $p < 0.001$, is shown. A pGL3-Enhancer vector, containing an SV40 promoter, was used as positive control in the assay (not shown). Values show means ± SD from six samples and represent ratio of relative light unit (RLU) of firefly luciferase to *Renilla* luciferase.

3.3. Cell line expression

In order to obtain a cell line endogenously expressing m*cmklr1*, northern blot analysis of 13 cell lines was performed using the m*cmklr1* ORF as probe (Fig. 3). A major transcript of 2.6 kb was detected in several cell lines, with the strongest expression in NB 4 1A3. The cell line, 3T3 clone A31, did not show any expression of the receptor gene.

3.4. Determination of the transcription start site

The transcriptional initiation site of m*cmklr1* was determined by 5′-RACE using cDNA from the NB4 1A3 cell line. 5′-RACE resulted in a 650-bp fragment (Fig. 4A), which was subcloned into pBluescript (SK-) for sequence analysis. Sequencing of six clones revealed four transcriptional start sites (Fig. 4B) to be located within the region of 255 to 275 nucleotides upstream of the translational start codon. We suggest that the start site located 275 nucleotides upstream of the translational start codon is the putative transcription start site (designated + 1 in Figs. 4B and 5A). This is the only site that starts with an adenine residue and shows high homology with the established initiator motif, 5′ PyPyA_{+1}N(T/A)PyPy, where Py = C or T and A_{+1} = the transcription start site (Lo and Smale, 1996). Other less well-defined start sites may represent various short versions generated during the 5′-RACE.

3.5. Characterization of the 5′ region of mcmklr1 and comparison to hcmklr1

The 5′ flanking region of m*cmklr1* was obtained by genome walking using BAC DNA as template. The amplified fragment was subcloned and sequenced (sequence data deposited in GenBank, accession no. AY 342406). In order to identify the promoter region, a 1039-nucleotide sequence upstream of the transcription start site was analyzed in search of transcription consensus motifs using the TRANSFAC database (Heinemeyer et al., 1999). The region did not contain any TATA box or consensus sites for CCAAT binding proteins, but several other transcription factor binding sites including several GC box-like motifs, AP-4 elements, C/EBP motifs, and GATA-1 and GATA-2 binding sites (Fig. 5A).

The 5′ flanking regions of h*cmklr1*a and *hcmklr1*b were found in GenBank (contig. NT 035235: 4570155–4571172). A sequence of 1017 nucleotides upstream of the transcription start site of h*cmklr1*a was analyzed for transcription binding elements. The putative human promoter region also lacks typical TATA and CCAAT boxes but contains several Sp1-, C/EBP-, GATA-1-, and GATA-2-binding sites (Fig. 5B), showing striking similarity with the mouse promoter.

3.6. Functional activity of the mcmklr1 promoter in NB4 1A3 cells

In order to functionally localize the promoter region regulating m*cmklr1* expression, 5′ deletions of a construct

Fig. 5. Nucleotide sequences of the 5′ flanking regions of the *cmklr1* gene, in mouse and human. (A) The mouse promoter was isolated from a BAC clone, subcloned, sequenced (accession no. AY342406) and analyzed in search of transcription consensus motifs. The predicted transcription start site is marked in bold (+1) and indicated by arrow, and the translation initiation codon is also shown in bold letters. Potential transcription binding motifs are presented in bold letters and are boxed. Sequences corresponding to Sp1 probes in EMSA experiments are underlined. (B) The human promoter sequence was obtained from GenBank (contig. NT 035235: 4570155–4571172). Transcription starts of the two human mRNA transcripts (accession no. U79526 for hCMKLR1a and U79527 for hCMKLR1b) are shown in bold and indicated by arrow. Potential transcription binding motifs are presented in bold letters and are boxed.

containing a 5′ flanking region of m*cmklr1* were performed generating four fragments of different length. The fragments were analyzed functionally in NB4 1A3 cells (Fig. 6) using the pGL3-Enhancer vector, containing a SV40 Enhancer element located downstream of the luciferase gene. It was found that the putative region promoting transcription is situated very close to the transcriptional start site. Thus, the construct with the shortest insert length (− 280 bp) showed the highest luciferase activity, while constructs of larger sizes gave successively decreasing activity. In comparison to the empty pGL3-Enhancer vector, the luciferase activity of the promoter constructs was generally low.

In order to test the tissue specificity of the promoter, the deletion constructs were transfected into mouse 3T3 clone A31 cells, a cell line that is not expressing m*cmklr1* endogenously. There were significant differences between the signal of the empty vector and that displayed by any of the promoter constructs. The fact that this cell line also

Fig. 7. Electrophoretic mobility-shift and super-shift assays using 1Sp1 (A), 2Sp1 (B), 3Sp1 (C) and 4Sp1 (D) labelled DNA probes. Mobility-shift assays were performed using a volume of NB4 1A3 nuclear extracts corresponding to 6 μg protein in each reaction. Free probe is indicated by an asterisk and gel shift with an arrow. In the competition assays, various excess levels (50–800 ×) of cold competitors were used (indicated as Fold × 10). Super-shift assays were carried out to identify the DNA-binding protein. Labelled probes were incubated with nuclear extracts in the presence or absence of Sp1 antibody (4 μg). The super-shift is indicated with a bracket.

showed luciferase activity when transfected with the promoter constructs indicated that the cell specificity does not reside within the 4-kb region analyzed. Thus, our analysis supports that the cloned region contains a functional promoter element but that the tissue-specific regulation appears to be dependent on additional elements.

3.7. Interactions between DNA-specific nuclear protein factors and promoter elements

The 280-bp fragment having the highest transcriptional activity was found to contain four GC- motifs composing binding sites for Sp1 binding proteins. The GC- motifs all shared a strong homology with the classical GC box, 5′GGGGCGGGG3′ (Bouwman and Philipsen, 2002). EMSA was performed to determine whether these tentative sites were actually occupied by DNA-binding proteins. The radiolabelled double-stranded oligonucleotides listed in Table 1 were incubated with nuclear extracts from NB41 A3 cells. The four Sp1 regions displayed similar mobility shift patterns, showing formation of similar-sized DNA–protein complexes, indicating that the same protein bound to all four probes.

Competitive binding assays were performed in order to confirm the specificity of the gel-shifts obtained. Depending on which individual probe was applied, competition of the binding to the protein using cold competitors showed a varying degree of effect. Using the labelled 1Sp1 probe, a gradually decreasing intensity of the gel-shift was seen when the reaction contained 50–100 × molar excess of unlabelled consensus Sp1 sequence or "self-probe" (1Sp1), and the shift was completely abolished at an 800 × molar excess level (Fig. 7A). When the 1Sp1 probe contained mutations of three nucleotides in the Sp1 binding site (Table 1) it was unable to compete for the binding to the protein. This indicated that the DNA–protein complex was specific.

For the 2Sp1 probe, 50 × molar excess of cold consensus Sp1 probe was enough to almost abolish the gel shift, while a slightly higher concentration of unlabelled self-competitor was needed to block the binding. Unlabelled 2Sp1 probe, containing the mutations in the Sp1 binding site, required an 800 × molar excess level to slightly affect the DNA–protein complex (Fig. 7B).

A 50 × molar higher concentration of cold consensus competitor completely abolished the resulting gel-shift with the labelled 3Sp1 probe. A slightly higher excess of cold self-probe was required to compete out the shift. On the other hand, 800 × higher concentration of unlabelled mutated 3Sp1 competitor was needed to block the obtained shift (Fig. 7C). Similar results were obtained with the 4Sp1 probe (Fig. 7D).

Super-shift assays, using polyclonal Sp1 antibodies mixed with probes and nuclear extracts, were performed to identify the DNA-binding protein. In the presence of Sp1 antibodies, all four DNA–protein complexes were super-shifted to a varying degree (Fig. 7A–D). Taken together, the results clearly identify the DNA-binding protein as the common transcription factor, Sp1.

4. Discussion

CMKLR1 is the provisional designation of an "orphan" GPCR, i.e., a receptor for which the natural ligand has not yet been identified. It shows sequence similarity to the subfamily of chemoattractant-type receptors. It was originally cloned and described in human (Gantz et al., 1996) and has also been presented under the name ChemR23 (Samson et al., 1998). Also, the mouse (DEZ or Gpcr27 in GenBank) and rat (CMKRL3) orthologues have been described (Methner et al., 1997; Owman et al., 1997). The fact that CMKLR1 can function as a coreceptor for HIV/SIV during viral entry into immune cells (Samson et al., 1998) warranted an exploration of the regulation of the receptor gene. As pointed out in the results two forms of the human receptor gene have been reported in GenBank (although no actual splicing data is available). There is only a two-amino acid difference in the coding region relevant for viral binding. Also, the receptor form used in the infection studies by Samson et al. (1998) is h*cmklr1b*, which suggests that the two extra amino acids are not required for viral coreceptor function. The mouse gene was chosen because it can also be used as a basis to develop transgenic animals in a further elucidation of receptor function.

Mapping of the mouse gene revealed high similarity to the human gene in its genomic organization, with a slight difference in the localization of the initial codon in the human splicing variant a. Using 5′-RACE it was possible to define the putative transcriptional initiation site, which shows high homology with the common initiator motif (Lo and Smale, 1996). Such motifs have been found in promoters both with and without TATA boxes, and are involved in the positioning of the constitutive machinery through binding to the basal transcription factor, TFII-D (Smale and Baltimore, 1989; Goodrich et al., 1996). The start site presently obtained differs from that reported by Methner et al. (1997) which may be a less likely start site in view of the absence of any initiator motif. The 2-bp difference in the sequence of this region may be due to genetic variability between the mouse strains used, if not simply a sequencing inaccuracy.

The 5′-region upstream of transcription start revealed that the mouse and the human genes have several additional features in common. Both lack typical TATA and CCAAT boxes, and both contain several putative Sp1-binding motifs (Suske, 1999), as well as AP-4 elements, C/EBP motifs, and GATA-1 and GATA-2 binding sites indicating that the genes are regulated in a similar manner. Since database search within the putative mouse promoter region suggested the presence of a multitude of possible transcription binding sites, a func-

tional study was conducted using deletion analysis. The results limited the regulatory activity to a GC-rich region located 280-bp upstream of the transcriptional start site. The overall transcriptional activity of this region was, however, low suggesting that auxiliary elements, such as enhancers necessary for augmenting trans-activation, were missing. The fact that the shortest deletion construct (c4) showed higher activity than the longer constructs (c1–c3) may be due to the presence of silencer elements within the longer segments.

It was possible to identify four potential Sp1-binding sites within the 280-bp region that are active in the transcription of the gene. This could be substantiated by the mobility shift/super-shift assays showing that all four motifs, indeed, were able to bind the Sp1 transcription factor. However, this does not necessarily imply that all four sites are involved in the functional regulation of the gene.

To conclude, the m*cmklr1* gene exhibits certain regulatory features in common with a housekeeping gene promoter, i.e., absence of a TATA box and the presence of several Sp1 binding sites (Kissonerghis et al., 1999). On the other hand, judging from the organization of h*cmklr1*, giving rise to two transcripts, the presence of additional exons within the mouse gene may be expected. It should be recalled that the widely expressed CXCR4, which is one of the two major coreceptors for HIV-1 in the viral infection process (Deng et al., 1996), also has features in common with a housekeeping protein. From a functional standpoint it is notable that mouse CMKLR1 has been shown to be specifically expressed during development of cartilage and osseous tissue (Methner et al., 1997), thus indicating that a specific regulation of the mouse gene does exist.

Acknowledgements

We thank Joanna Daszkiewicz-Nilsson and Margareta Pusch for expert technical assistance and Mikael Sigvardsson for critically reading the manuscript. This work was supported by the Foundation for Strategic Research (SSF, Inflammation Research program), the Swedish Research Council (Project No. 05680), GS Development, IngaBritt and Arne Lundberg's Foundation, Alfred Österlund's Foundation, the Crafoord Foundation, and the Royal Physiographic Society.

References

Andrews, N.C., Faller, D.V., 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res. 19, 2499.

Bouwman, P., Philipsen, S., 2002. Regulation of the activity of Sp1-related transcription factors. Mol. Cell. Endocrinol. 195, 27–38.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal. Biochem. 162, 156–159.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., Davis, C.B., Peiper, S.C., Schall, T.J., Littman, D.R., Landau, N.R., 1996. Identification of a major co-receptor for primary isolates of HIV-1. Nature 381, 661–666.

Gantz, I., Konda, Y., Yang, Y.K., Miller, D.E., Dierick, H.A., Yamada, T., 1996. Molecular cloning of a novel receptor (CMKLR1) with homology to the chemotactic factor receptors. Cytogenet. Cell Genet. 74, 286–290.

Goodrich, J.A., Cutler, G., Tjian, R., 1996. Contacts in context: promoter specificity and macromolecular interactions in transcription. Cell 84, 825–830.

Heinemeyer, T., Chen, X., Karas, H., Kel, A.E., Kel, O.V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F., Wingender, E., 1999. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. Nucleic Acids Res. 27, 318–322.

Henikoff, S., 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351–359.

Kissonerghis, M., Daly, G., Feldmann, M., Chernajovsky, Y., 1999. The murine p75 TNF receptor promoter region: DNA sequence and characterization of a cis-acting silencer. Mol. Immunol. 36, 125–134.

Lo, K., Smale, S.T., 1996. Generality of a functional initiator consensus sequence. Gene 182, 13–22.

Methner, A., Hermey, G., Schinke, B., Hermans-Borgmeyer, I., 1997. A novel G protein-coupled receptor with homology to neuropeptide and chemoattractant receptors expressed during bone development. Biochem. Biophys. Res. Commun. 233, 336–342.

Murphy, P.M., 1994. The molecular biology of leukocyte chemoattractant receptors. Annu. Rev. Immunol. 12, 593–633.

Owman, C., Lolait, S.J., Santén, S., Olde, B., 1997. Molecular cloning and tissue distribution of cDNA encoding a novel chemoattractant-like receptor. Biochem. Biophys. Res. Commun. 241, 390–394.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview, NY.

Samson, M., Edinger, A.L., Stordeur, P., Rucker, J., Verhasselt, V., Sharron, M., Govaerts, C., Mollereau, C., Vassart, G., Doms, R.W., Parmentier, M., 1998. ChemR23, a putative chemoattractant receptor, is expressed in monocyte-derived dendritic cells and macrophages and is a coreceptor for SIV and some primary HIV-1 strains. Eur. J. Immunol. 28, 1689–1700.

Smale, S.T., Baltimore, D., 1989. The "initiator" as a transcription control element. Cell 57, 103–113.

Suske, G., 1999. The Sp-family of transcription factors. Gene 238, 291–300.

Watson, S., Arkinstall, S., 1994. The G-protein Linked Receptor Facts Book. Academic press, London.