The Human CC Chemokine Receptor 5 (CCR5) Gene

MULTIPLE TRANSCRIPTS WITH 5'-END HETEROGENEITY, DUAL PROMOTER USAGE, AND EVIDENCE FOR POLYMORPHISMS WITHIN THE REGULATORY REGIONS AND NONCODING EXONS

(Received for publication, August 12, 1997, and in revised form, September 29, 1997)

Srinivas Mummid, Seema S. Ahuja, Brent L. McDaniel, and Sunil K. Ahuja†

From the Department of Medicine, University of Texas Health Science Center at San Antonio and South Texas Veterans Health Care System, Audie L. Murphy Division, San Antonio, Texas 78284-7870

Human CC chemokine receptor 5 (CCR5), mediates the activation of cells by the chemokines macrophage inflammatory protein-1α, macrophage inflammatory protein-1β, and RANTES, and serves as a fusion cofactor for macrophage-tropic strains of human immunodeficiency virus type 1. To understand the molecular mechanisms that regulate human CCR5 gene expression, we initiated studies to determine its genomic and mRNA organization. Previous studies have identified a single CCR5 mRNA isoform whose open reading frame is intronless. We now report the following novel findings. 1) Complex alternative splicing and multiple transcription start sites give rise to several distinct CCR5 transcripts that differ in their 5'-untranslated regions (UTR). 2) The gene is organized into four exons and two introns. Exons 2 and 3 are not interrupted by an intron. Exon 4 and portions of exon 3 are shared by all isoforms. Exon 4 contains the open reading frame, 11 nucleotides of the 5'-UTR and the complete 3'-UTR. 3) The transcripts appear to be initiated from two distinct promoters: an upstream promoter (P1), upstream of exon 1, and a downstream promoter (P2), that includes the "intronic" region between exons 1 and 2. 4) P1 and P2 lacked the canonical TATA or CAAT motifs, and are AT-rich. 5) P2 demonstrated strong constitutive promoter activity, whereas P1 was a weak promoter in all three leukocyte cell environments tested (THP-1, Jurkat, and K562). 6) We provide evidence for polymorphisms in the noncoding sequences, including the regulatory regions and 5'-UTRs. The structure of CCR5 was strikingly reminiscent of the overall structure of other chemokine/chemoattractant receptors, underscoring an important evolutionarily conserved function for a prototypical gene structure. This is the first description of functional promoters for any CC chemokine receptor gene, and we speculate that the complex pattern of splicing events and dual promoter usage may function as a versatile mechanism to create diversity and flexibility in the regulation of CCR5 expression.

CCC chemokine receptor 5 (CCR5), a receptor for the CC chemokines macrophage inflammatory protein-1α, macrophage inflammatory protein-1β, and RANTES (1-3), also serves as a fusion cofactor for the entry of macrophage-tropic strains of HIV-1 (4-8). The level of CCR5 cell surface expression may have a direct influence on the relative ease with which an individual acquires HIV-1 infection (9-12): individuals homozygous for a 32-bp deletion (denoted ΔCCR5) in the open reading frame (ORF) do not express the protein on the cell surface, and are relatively resistant to developing HIV-1 infection. In contrast, individuals who display the CCR5/ΔCCR5 genotype can develop HIV-1 infection, however, their progression to AIDS may be slower. Interestingly, in individuals who display the CCR5/CCR5 genotype, the cell surface expression of CCR5 can be highly variable (13), however, whether this heterogeneity in protein expression also correlates with differences in HIV-1 infection/transmission in vivo is not known. These observations suggest that a therapeutic or preventive strategy based on targeting CCR5 cell surface expression could potentially be quite beneficial. Toward this end, we have initiated studies to define the structural organization of CCR5 and molecular factors that regulate its expression.

Phylogenetic analysis of the G-protein coupled receptor (GPCR) superfamiliy indicates that replication of a progenitor gene may have given rise to clusters of evolutionarily related receptor genes (14, 15). Two such GPCR clusters are members of the chemokine receptor subclass, and receptors for the classical chemoattractants, such as the N-formyl peptide receptor. To date, the complete mRNA and genomic organization of only a limited number of human chemokine receptors has been described (16-20), however, a comparison of their structural organization with that of the receptors for the classical chemottractants reveals some striking similarities (21-24). 1) Their ORFs are usually intronless or contain a single intron interrupting the amino-terminal coding region, as is the case for the C5a receptor (21). 2) Their 5'-untranslated regions (UTR) can have a surprisingly complex organization. Unlike most GPCRs, the 5'-UTRs for these genes reside on multiple exons and alternative splicing may generate multiple mRNA isoforms. 3) Splicing of the untranslated exons to form the mature transcripts occurs at a common 3'-splice junction that is a short distance upstream of the start of the translation. Thus, the transcription and translation start sites can be separated by long intervening sequences. 5) Although they are products of

† To whom correspondence should be addressed: Dept. of Medicine, Div. of Infectious Diseases, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr., San Antonio, TX 78284-7870. Tel.: 210-567-6511; Fax: 210-567-4654; E-mail: ahujas@uthscsa.edu.

‡ This work was supported by a Kleberg Foundation Grant (to S. S. A. and S. K. A.), a Howard Hughes Medical Research Institute New Faculty Award and American Heart Association, Texas Affiliate Grant (to S. S. A.), a Howard Hughes Medical Research Institute New Faculty Award and American Heart Association, Texas Affiliate Grant. This work was supported by a Howard Hughes Medical Research Institute New Faculty Award and American Heart Association, Texas Affiliate Grant. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s) AF031236 and AF031237.

1 The abbreviations used are: CCR, CC chemokine receptor; ORF, open reading frame; UTR, untranslated region; GPCR, G-protein coupled receptor; bp, base pairs; kb, kilobases; 5'-RACE, rapid amplification of cDNA end; RT-PCR, reverse transcriptase-polymerase chain reaction; DARC, Duffy antigen receptor for chemokines; HIV, human immunodeficiency virus; TNF, tumor necrosis factor; CMV, cytomegalovirus.
distinct genes, they tend to be physically clustered on a single chromosome (14, 15, 19, 21, 25). For example, CCR5 and several other CCRs co-localize on chromosome 3p21.3-p24 (25), whereas several of the chemotaxtractant receptors co-localize to 19q13.3 (21). These similarities suggest that despite coding for receptors that have diverse ligand-receptor relationships, these two subclasses of receptors have retained a remarkably conserved structural organization.

Some of these prototypal structural features also appears to be true for human CCR5. First, a partial length gene (1376 bp) has been cloned and it has an intronless CCR5 ORF (Ref. 1; position 240 to 1298). Second, cDNA clones for CCR5 have been cloned and reported by two groups (2, 3). Comparison of the partial CCR5 sequence with that of the cDNA clones, and restriction mapping of P1 clones suggests the presence of a single ~1.9 kb intron between position ~11 and ~12 relative to the start of translation (1–3). To delineate the full extent of the 5′-UTR of human CCR5, Rapport et al. (3) also performed 5′-RACE (5′-rapid amplification of cDNA ends) on human spleen cDNA, and by this method the longest 5′-UTR identified was 54 nucleotides in length. The CDNA clone reported by Rapport et al. (3) also contains a poly(A) tail, suggesting a full-length 3′-end. Nevertheless, the exact location of the remainder of the reported CCR5 5′-UTR sequence on the gene, and the nature of the cis-acting elements is not known.

Expression of CCR5 at the mRNA level suggests that CCR5 may contain tissue-specific cis-acting elements. An ~4-kb human CCR5 transcript has been observed in several human cell lines, and in human thymus, spleen, small intestine, and peripheral blood leukocytes (1–4). Combadiere et al. (2) have shown that human CCR5 transcripts are present in primary adherent monocytes but are absent from the primary neutrophils and eosinophils. Carroll et al. (26) have reported recently that human unstimulated CD4+ cells do not express CCR5 mRNA. However, CD4+ cells activated by phytohemagglutinin (PHA)/IL-2 expressed CCR5 mRNA, whereas those costimulated with immobilized antibodies to CD3/CD28 did not. Both unstimulated CD4+ cells and CD4+ cells costimulated with CD3/CD28 were resistant to infection by macrophage-tropic strains of HIV-1 in vitro, whereas phytohemagglutinin/interleukin-2 activated CD4+ cells could be infected, further highlighting the importance of understanding the molecular mechanisms that regulate CCR5 expression.

The present study represents a first step toward addressing some of the issues outlined above. Unlike reported previously, we demonstrate that the mRNA structure of human CCR5 is not monomorphic. Instead transcript analysis by 5′-RACE and RT-PCR (reverse transcriptase-polymerase chain reaction) reveled complex alternative splicing patterns in the 5′-UTRs of CCR5: alternative splicing of four exons that span ~6 kb of CCR5 give rise to multiple CCR5 transcripts that differ in their 5′-UTRs. Although the generation of multiple CCR5 transcripts has no effect on the protein sequence of CCR5, it does have consequences for the regulation of the gene as we demonstrate that CCR5 transcription is regulated by at least two promoters, and we ascribe an important role for the 5′-UTR and intron sequences in regulating CCR5 expression. In this report we also provide evidence that the regulatory sequences and noncoding exons of CCR5 are polymorphic.

**EXPERIMENTAL PROCEDURES**

**Cells and Cell Culture**—After obtaining informed consent, normal adult donors were pretreated with granulocyte colony stimulating factor (Amgen; 10 μg/kg body weight, subcutaneously) for 5 days, and then their low density cells in the peripheral blood were collected by apheresis. These cells were enriched for CD34+ progenitor cells by positive selection, using the Cepatre SC column (CellPro, Bothell, WA). The purified CD34+ cells were differentiated into dendritic cells by cultur-
**Human CCR5 Gene Regulation**

**RESULTS**

**Heterogeneity in the 5'-UTR of Human CCR5 mRNA**—A single CCR5 mRNA isoform that contains a 5'-UTR of 54-nucleotides in length has been reported (3). Since alternative splicing in the 5'-UTRs appears to be a feature common to several human chemokine and chemoattractant receptors (18, 22, 24), we hypothesized that this might also be true for CCR5. To test this hypothesis, we designed a strategy that involved 5'-RACE and RT-PCR techniques, and probed the diversity in the CCR5 mRNA structure in several primary human cell types and the human cell lines THP-1 and Jurkat. By this strategy, we identified PCR products of ~100 to ~350 bp in length from human dendritic cells, suggesting the possibility of novel 5'-UTR sequences. These PCR products were subcloned, and a schematic illustration of their sequence composition is shown in Fig. 1A. Based on sequence analysis and criteria outlined below, we have arbitrarily segregated these cDNA clones into two categories, representing either “full-length” or “truncated” CCR5 transcripts.

The two full-length CCR5 transcripts, designated as CCR5A and CCR5B, shared three sequence segments but differed by the presence or absence of a 235-bp sequence segment in the 5'-UTR (Fig. 1). As demonstrated later, these sequence segments were identified on CCR5, and based on their location on the gene they were designated as exons 1–4; exon 2 corresponded to the 235-bp sequence segment that is unique to...
CCR5A. Exons 1, 3, and 4 were common to both CCR5-A and -B, and the ORF, 11-bp of the 5’-UTR and the 3’-UTR resided in exon 4 (Fig. 1A).

The cDNA clones that lacked sequences corresponding to the 5’-most unique sequence segment, i.e. exon 1, were arbitrarily classified as truncated CCR5 mRNA isoforms. The 5’-termini of the truncated clones relative to their position on CCR5A are shown in Fig. 1. It should be emphasized that the truncated CCR5 transcripts could also represent incomplete cDNA synthesis by the reverse transcriptase. However, two findings suggest that this may not be the case. 1) From a single RT-PCR, we cloned products whose lengths were significantly longer than the truncated transcripts (Fig. 1A). Except in a single instance, several clones had identical 5’ termini, suggesting that they may represent transcripts that originate from distinct transcription start sites. It should also be noted that the presence of additional CCR5 isoforms that may have unique 5’ noncoding exons or novel splice patterns cannot be excluded.

The cDNA sequence reported by Raport et al. (3) lacked in-frame stop codons in the 5’-UTR, raising the possibility of a longer CCR5 ORF initiated at an upstream methionine. In-frame stop codons were identified 26 and 12 amino acids upstream of the currently assigned translation initiation codon in CCR5A and CCR5B, respectively. None of the upstream in-frame amino acids were a methionine, excluding the possibility of a longer transcript that could encode a protein isoform with an amino-terminal extension. Interestingly, four upstream AUG triplets were found in the 5’-UTR of both CCR5A and CCR5B, but they were followed by downstream termination codons, and the two longest minicistrons were 9 and 15 amino acids in length.

The 5’-UTR sequences of the full-length and truncated CCR5 transcripts appeared to be highly conserved in evolution as GenBank data base analysis revealed strong sequence homology with the 5’-UTRs of mouse and rat CCR5 cDNAs (Fig. 1B). The 5’ termini of the 5’-UTRs of mouse and rat cDNAs reside in a region that corresponds to exon 2 of human CCR5A. Whether additional upstream mRNA sequences exist in these two species is not known. It is interesting that 12 bp upstream of the start of the translation start site, all the human CCR5 cDNA clones had a 4-bp insertion (CCCC) relative to the mouse and rat cDNAs.

**Tissue Distribution of Human CCR5 mRNA Isoforms**—All the CCR5 cDNA clones identified contained exon 4 and portions of exon 3, and the additional length contributed by exons 1 and/or 2 to CCR5A or CCR5B was not substantial. This implied two points. First, that the proportion of transcripts in human cell types that are either full-length or truncated cannot be readily ascertained by size differences on Northern blots. Second, since CCR5A and CCR5B can be differentiated only by the presence or absence of exon 2, a RT-PCR strategy could be designed to evaluate exon usage in different human leukocyte populations. However, the latter strategy would not be helpful in defining the relative abundance of the truncated transcripts, as portions of exon 3 are common to all isoforms. To illustrate the first point, when we used a probe that corresponded to exon 1, a ~4.0-kb hybridizing band was visualized in human poly(A)+ mRNA derived from bone marrow, peripheral blood mononuclear cells, thymus, lymph node, and spleen (data not shown), and corresponded to the transcript size seen in the identical tissues hybridized with an ORF/3’-UTR probe (3).

The second point is illustrated in Fig. 2, which demonstrates the splicing patterns, i.e. exon usage, of CCR5 mRNA. In these RT-PCR experiments, total RNA derived from the primary human cell types shown in Fig. 2, and the THP-1 and Jurkat cell lines were used as PCR templates (location of the final set of RT-PCR primers is shown in Fig. 1B). In these experiments, we observed two bands in these cell types (Fig. 2, and data not shown for THP-1 and Jurkat cell lines). To confirm the exon composition of the ethidium bromide-stained PCR products, we subcloned the two bands that were amplified from dendritic cells and the THP-1 cell line. Sequence analysis revealed that the upper and lower band in Fig. 2 corresponded to isoforms that contained exons 1 + 2 + 3 (CCR5A) or exons 1 + 3 (CCR5B), respectively. It should be noted that this analysis is qualitative, and although minor variations in the proportion of the transcripts containing these exons were observed, there was no clear pattern of tissue-specific utilization of either CCR5A or CCR5B.

**The Human CCR5 Gene**—Using PCR we amplified, cloned, and sequenced overlapping fragments of human CCR5, that together comprised an ~8-kb contiguous stretch of CCR5. The 5’-UTR sequences detected by 5’-RACE and RT-PCR, and the cDNA sequence reported by Raport et al. (3) were identified on this genomic contig (Fig. 3A). This genomic contig spanned 8035 bp, and originated ~1.9 kb upstream of exon 1 and terminated immediately upstream of the polyadenylation signal (Fig. 3A). The gene is organized into four exons and two introns (Figs. 3A and 4A). Both introns interrupt the 5’-UTR. Interestingly, exons 2 and 3 are contiguous and are not interrupted by an intron. The exon/intron splice junctions in CCR5 conform to the consensus sequences for 5’-CAAGTGATAG and 3’-Y(NY) splice sites (the invariant dinucleotides at the termini of the intron consensus sequences are underlined; Fig. 3A). Interestingly, a region upstream of exon 1 had strong sequence homology (~89%) with sequences in the 3’-flanking region of CCR5 (Fig. 3B).

We compared the 5’- and 3’-flanking regions of CCR5 with sequences deposited in GenBank (updated July 31, 1997). This analysis revealed identity or close homology between the CCR5 sequences that we characterized and two unpublished gene sequences that were submitted while this work was in progress. 1) The entire 8055-bp sequence that we cloned was colinear with a portion of a human genomic DNA contig sequenced as part of the Advanced Genome Sequence Analysis Course, Cold Spring Harbor Laboratory, NY (GenBank accession number U95626); this unpublished contig is 143,068 bp in length and in addition to CCR5, it contains CCR2A, CCR2B and an orphan chemokine receptor gene. Our CCR5 sequence ends just proximal to the polyadenylation signal. However, alignment of our sequence contig with the sequences contained in GenBank accession U95626 revealed that the nucleotides that follow the end of our clone are identical to the polyadenylation

![FIG. 2. Expression of CCR5A and CCR5B in human leukocytes.](image-url)
signal sequence AAATAA. 2) A 227-bp sequence that is upstream of the Macaca mulatta CCR5 ORF (GenBank accession number U77672) had a high degree of homology with the region that corresponds to intron 2 of human CCR5. The 5' and 3'-flanking sequences reported previously by Samson et al. (1) were 239- and 78-bp in length, respectively, and identical sequences were found in the CCR5 that we characterized. A region in intron 2 also had strong sequence homology with Alu repeats (Fig. 4).

The exact location of the exon/intron boundary between intron 2 and exon 4 in human CCR5 appears to be conserved in mouse. Comparison of the mouse CCR5 cDNA and genomic sequences (GenBank accession numbers D83648 and U68565) revealed an intron between -11 and -12 upstream of the translation start codon, a position that is identical for intron 2 in the human CCR5. Interestingly, the 5'-UTRs of human and mouse CCR5 are highly conserved (Fig. 1).

Evolutionary Conservation in the mRNA and Genomic Structure of Human CCR5 with That of Other Human Chemokine/Chemoattractant Receptors—

FIG. 3. Human CCR5 sequence. A, exon sequence is in uppercase, introns and the 5'-flanking region are in lowercase. ORF sequences are shown in uppercase boldface letters and the derived amino acids are indicated by a single-letter code below the first nucleotide of each codon. The asterisk denotes the stop codon. Exons and intron names are left justified above their 5' termini. Double underline indicates terminal dinucleotides for the introns defined by each of the CCR5 RNA splice variant shown in Fig. 1. Note that all these dinucleotides obey the GT/AG rule and in one instance the AG dinucleotide resides within the exon defined by another splice variant. The 5'-most expressed nucleotide has arbitrarily been designated as nucleotide +1 (see also Fig. 1B). Gaps are denoted by serial dots. The length of the sequence not shown is indicated in the gaps. The pyrimidine-rich sequences are overlined by a straight line. Sequences that conform favorably to the indicated transcription factor DNA binding elements are overlined by a straight arrow: the direction of the arrow indicates the 5' → 3' orientation of the putative binding site. The inverted L-shaped arrows delimit a region with strong homology (in reverse orientation) to a region in the 3'-UTR of CCR5 (panel B). Uppercase italicized sequence in the 3'-end of CCR5 represents the polyadenylation signal sequence (AAATAA). Alignment with the sequences contained in GenBank accession number U95626 revealed that this signal sequence is immediately contiguous with the CCR5 contig that we cloned. B, sequence alignment depicting the high degree of homology between a short region in the 5'- and 3'-flanking sequences of CCR5. Note that the 3'-flanking sequence is in the reverse complement orientation.
tors (16–18, 20, 22, 24), suggesting a selective evolutionary pressure for these receptors to retain a conserved gene architecture (Fig. 4). It should be appreciated that, to date, the gene and mRNA structures (human) of only one CCR, CCR2 (20), two CXCRs, CXCR1 and CXCR2 (18, 19), and the Duffy antigen receptor for chemokines (DARC; 16, 17) has been described. Furthermore, the functional promoters for only two human chemokine receptors, CXCR1 and CXCR2, have been described (18). As described below, we have characterized two promoters for CCR5, designated as Pu and Pd, and their locations are noted in Fig. 4. Interestingly, as is the case for the promoters for CXCR2 (18) and platelet-activating factor receptor gene (22, 23), the two CCR5 promoters are also tandemly arranged on the gene. Another feature that is common to both CCR5 and CXCR2 is that they contain exon-exon units that are uninterrupted by an intron. For example, exon 2 of CCR5A, resides in the “intronic” region for CCR5B, and exon 5 of the CXCR2–3 isoform, resides in the intronic region for CXCR2–1, -2, and -4 isoforms (Fig. 4).

**Molecular Dissection of Functional Promoters for CCR5—**

The genomic region upstream of exon 1 should potentially contain the cis-acting elements important in the promoter activity of CCR5A and CCR5B. We therefore constructed CCR5-firefly luciferase chimeric plasmids from portions of the gene upstream of exon 1, designated as pA1–4 (Fig. 5, upper panel). We tested the ability of these promoter constructs to drive the expression of the reporter gene (firefly luciferase) in the following cell lines: 1) THP-1, a human monocytic leukemia cell line, a surrogate for monocytes; 2) K562, a human chronic myelogenous leukemia cell line, a surrogate for undifferentiated he-
mopoietic cells; and 3) Jurkat, which is a human T cell leukemia cell line. To correct for differences in transfection efficiency, we co-transfected the promoter constructs and the promoterless vector pGL3-Basic with pRL-CMV, a construct that contains the renilla luciferase gene downstream of a CMV promoter. Lysates prepared from cells transfected with constructs pA1–4 exhibited weak luciferase activity (Fig. 5). This genomic region upstream of exon 1, which has weak promoter activity, is designated as the upstream promoter (PU).

Because a large number of 5′-RACE clones terminated either in exon 3 or at the 3′-end of exon 2 (Fig. 1A), we hypothesized that these transcripts may represent distinct isoforms that are initiated because of the usage of an alternative promoter. To test this hypothesis, we constructed the series of promoter constructs shown in the lower panel of Fig. 5. It should be noted that in some instances these constructs contain portions of PU, intron 1, and exon 2, and that the distal end of each of these constructs resides within exon 3.

In contrast to PU, the region upstream of exon 3, designated as the downstream promoter (PD), had strong luciferase activity in all the three cell lines tested (Fig. 5, lower panel). Maximal promoter activity was consistently observed in the cell lysates from K562 cells, especially with those transfected with pB3 and pB4. The promoter activity for these two constructs in K562 cells was ~8–10-fold more than that detected in cells transfected with pB1, pB2, or pB5. The increase in luciferase activity in THP-1 and Jurkat cell lines transfected with pB3 and pB4 was not as prominent as that observed for these two promoter constructs in K562 cells. Relative to PB3 and PB4, the construct pB5 exhibited weak promoter activity. This finding suggests that the sequences between pB4 and pB5 may contain important cis-acting elements for CCR5 promoter activity. It is important to note that since all the PD constructs contain all or portions of exon 2, it is likely that this noncoding exon may play an important role in modulating gene expression.

Analysis of the PU and PD Sequences—It is important to appreciate that because of the complex genomic and mRNA organization of CCR5, it is difficult to unambiguously assign certain regions of CCR5 as an exon, intron, or promoter. Notwithstanding this caveat, PU and PD lacked canonical TATA and CCAAT motifs. However, in PD there was a nonconsensus TATA box (TTTTA; Fig. 3A). Unlike most TATA-less promoters which have a high GC content, PU and PD were GC poor. The overall G + C content of PU and PD was ~46 and ~40%, respectively. We identified several pyrimidine-rich segments in both PU and PD. Pyrimidine-rich sequences have been observed in the proposed promoter for DARC (17), and several other genes that are abundantly expressed in myeloid cells, including N-formyl peptide receptor (FPR) (24). PU and PD contained consensus sequences for several transcription factor DNA-binding sites (e.g., AP-1, Oct-1, PuF, PU.1, and NF-xB-like). The PU.1 transcription factor has been found to be important in the promoter activity of several genes expressed in myeloid cells, including monococyte colony-stimulating factor and CD11b genes (31). Multiple binding sites for GATA-1, an important transcription factor in the development of hematopoietic cells (31), and for Sp1 were also noted.

Polymorphisms in CCR5 Noncoding Sequences—We aligned the nucleotide sequences of the CCR5 gene that we cloned with gene sequences in GenBank accession number U95626, and the sequences of the cDNA clones derived by RT-PCR and 5′-RACE. This alignment revealed extensive nucleotide differences in the noncoding sequences of the gene. The relative positions of the nucleotide substitutions, deletions, or inser-
Fig. 6. Differences in the noncoding sequences of CCR5. The alignment of the following nucleotide sequences is shown. 1) CCR5 gene isolated in this report (top line); 2) the sequences in GenBank (accession number U95626) that are co-linear with the sequences determined in this report; 3) partial CCR5B cDNA clone (source: dendritic cells) from a normal donor; 4) partial CCR5A cDNA clone (source: dendritic cells) from an unrelated second donor; 5) partial CCR5A cDNA clone from THP-1 cells. The nucleotide numbers are derived from Fig. 3A. Serial dots denote gaps introduced. Boxes, denote insertions or deletions. PU denotes upstream promoter.

From an evolutionary perspective, it is intriguing that in addition to their ORFs, the 5'-UTRs of mouse, rat, and human CCR5 share strong sequence homology. To date, murine homologues for CCR1–5 have been cloned (reviewed in Ref. 35). The 5'-UTR sequences for murine CCR1 are not available in GenBank, nevertheless, unlike the strong interspecies homology of the 5'-UTRs of CCR5 (Fig. 1B), the 5'-UTRs of mouse and human CCR2, CCR3, and CCR4 do not share significant sequence homology (data not shown). These observations point toward a selective pressure for both mouse and human CCR5 to retain similar noncoding exons, which at least in humans, may participate in CCR5 gene regulation.

It is likely that CCR5 regulation may occur at many levels (14, 15). As is the case for other GPCRs, the cell surface expression of CCR5 may be regulated at the protein level, over the short term, through mechanisms such as receptor internalization, sequestration, and desensitization. Longer term regulation of these receptors is likely to be achieved through regulation of the rate of transcription of the gene, stability of the mRNA, and translation efficiency, and there is increasing evidence that the sequences in the 5'- and 3'-UTRs may influence these processes (36).

We will discuss two possible mechanisms by which the 5'-UTRs of CCR5 may regulate gene expression. First, the 5'-UTR of CCR5-A and -B have several structural features that may exert a negative effect on the efficiency of translation. Kozak has examined factors in the 5'-UTRs that promote efficient translation (37, 38), which include the observation that: 1) most eukaryotic mRNAs have a short 5'-UTR, and 2) there are no AUGs upstream of the translation initiation site of the major ORF. Both CCR5A and CCR5B, the two full-length transcripts, have relatively long 5'-UTRs, and they belong to the unusual class of mRNAs (<10% vertebrate RNAs characterized) that contains AUG triplets upstream of the AUG that initiates the major ORF (Fig. 1B). The presence of translation initiation codons followed immediately by termination codons creates short upstream ORFs in the 5'-UTR. As reported in other gene systems (39, 40) these short upstream ORFs could lead to reduced protein output through a mechanism of abortive translation. For example, a product of a short upstream ORF encoding a 19-amino acid leader peptide inhibits the translation of the β2-adrenergic receptor (40). Since some of the truncated isoforms lack short upstream ORFs, it is conceivable that preferential initiation of transcripts from P1 may represent a potential mechanism by which CCR5 expression is modulated, as this would by-pass the possible inhibitory effects of the upstream minicistrons.

A second mechanism includes the possibility that differences in the secondary structures of the 5'-UTRs of the distinct CCR5 transcripts may influence translation efficiency. It is known that a Gibbs free energy of formation (ΔG) of less than −50
kcal/mol can impair the passage of the ribosomal 40 S subunits as they scan from the cap site (41). Algorithms developed by Dr. M. Zucker (Ref. 42 and http://www.ibc.wustl.edu/~zucker/rna) were used to analyze the 5′-UTRs of CCR5A and CCR5B for their tendency to undergo secondary structure. These algorithms predict that the ΔG of CCR5A and CCR5B are −69.5 and −48.7 kcal/mol, respectively, suggesting that relative to CCR5B, CCR5A has a higher propensity to form a very stable structure.

We have identified two CCR5 promoter regions that were active in all three cellular environments tested: P_U, a weak promoter which resides proximal to exon 1, and P_D, a stronger promoter which is located upstream of exon 3. It is conceivable that regions further upstream of exon 1, or constructs shorter than those that we tested, may support strong promoter activity for P_U. We ascribe an important role for the region between +429 to +634 in regulating CCR5 expression. Although within this region, consensus sequences representing binding sites for transcription factors such as Oct-1 and GR-β are present, future studies will be required to determine the precise cis-acting elements that confer this activity. It should be noted, that several of the constructs designed to test P_D had intron 1 and exon 2 sequences, implicating an important function for these two regions in the regulation of CCR5. An important role for intronic sequences in the regulation of several genes has been described, including for CXCR2 (18).

The promoter sequences of CCR5 have two interesting features. First, a region in P_U has sequence homology to a region in the 3′-UTR, the significance of which, if any, remains unclear. Second, characteristic of several GPCRs, neither P_U nor P_D has classical TATA or CCAAT motifs, although P_D does contain a nonconsensus TATAA box. Most genes that are TATA-deficient can be divided into two classes on the basis of their upstream GC content (43). GC-rich promoters, found primarily in housekeeping genes, are very complex and prevalent; their promoters contain several binding sites for the ubiquitous trans-activating Sp1 protein and have several transcription start sites. In contrast, the remainder of the genes that are TATA-deficient and are not GC rich, tend to be regulated during differentiation or development; many of their promoters are not constitutively active and initiate at only one or a few very tightly clustered start sites. The AT-rich composition of the CCR5 promoters, P_U or P_D, suggests that they belong to the latter class of promoters. However, in contrast to this subclass of TATA-deficient promoters, P_U or P_D appear to be constitutively active, are possibly initiated at several transcription start sites, and there is no conclusive evidence, to date, to suggest that CCR5 requires strict activation and inactivation during cellular differentiation and development.

It is clear from the study of several diverse gene systems that alternative promoter usage resulting in alternative transcripts is an important evolutionary mechanism to create diversity in the regulatory control of gene expression (reviewed in Ref. 32). In these systems, alternative promoter usage has been shown to be an important transcriptional mechanism for regulating either tissue- or cell-type specific expression, the level of expression, the developmental stage-specific (temporal) expression, the specific capacity to respond to a particular cellular or metabolic conditions, or the translational efficiency of the mRNA. Which, if any of these mechanisms is operative in CCR5 is not known, however, several possible scenarios for CCR5 can be envisaged. It is possible that the level of CCR5 expression is regulated at a transcriptional level by the usage of promoters of different strengths, such as the ones we have described. Whether the observed differences in transcription efficiencies between the P_U and P_D, and whether the observed differences in the transcription efficiencies in the P_D constructs in the different cell environments might result in differential expression at the protein level is unknown.

Although the protein encoded by the different CCR5 transcripts is likely to be identical in different cell types, they may be regulated differentially in these different cell types by various extracellular signals, such as cytokines or chemokines. To test this latter possibility, we have in preliminary experiments, determined whether cytokine stimulation alters the constitutive promoter activity of a single promoter construct (pB3). The promoter activity of pB3 in Jurkat cells stimulated with phytohemagglutinin, phytohemagglutinin and phorbol myristic acid, iomycin and phorbol myristic acid, or CD3/CD28 was similar to that observed in unstimulated Jurkat cells transfected with pB3 (n = 3; data not shown). Similarly, the cell lysates of THP-1 cells transfected with pB3 and stimulated with lipopolysaccharide, TNF-α, interleukin-6, and interferon-γ exhibited promoter activities similar to the cell lysates from the unstimulated THP-1 cells transfected with pB3 (n = 3; data not shown). Additional studies will be required to investigate whether these stimuli alter the promoter activity of the other reporter constructs used in this study.

Several polymorphisms have been described in the CCR5 ORF (10–12, 44). We now provide evidence for polymorphisms in the flanking regions of CCR5, however, it should be emphasized that the significance of this finding, if any, in HIV-1 infection is not known. Several studies have clearly demonstrated that genes can be polymorphic not only in their coding regions, but also in important cis-regulatory sequences (45–53). Furthermore, transcriptional mutants may profoundly affect the promoter strengths of particular alleles by altering the affinity of regulatory proteins for these elements, and in some instances a single nucleotide change in a critical regulatory region can result in up to 1 order of magnitude difference in transcriptional activity of two otherwise identical promoters. As discussed below, this in turn, can have a profound affect on protein synthesis.

One of the most striking examples of transcriptional mutants affecting protein synthesis came in the wake of the cloning of the human β-globin gene nearly 20 years ago, where in addition to mutations in the coding region, single mutations in the regulatory regions were shown to decrease the amount of β-globin produced by red cells, leading to the blood disorder called β-thalassemia (52). It is interesting, that to date, over 300 β-thalassemia alleles have been discovered, including 12 transcriptional mutants, that account for the molecular basis of the marked heterogeneity of the β-thalassemia syndrome. Transcriptional mutants that lead to an increase in protein expression have also been described. For example, studies have linked the variant allele for the TNF-α gene, referred to as TNF-2, to increased serum levels of TNF-α, and a poor prognosis for several infections, such as malaria (53). Thus, it is conceivable that the polymorphisms in the regulatory regions of CCR5 may, in part, explain the observed variability in CCR5 expression in individuals who display the CCR5/CCR5 genotype (13, 54), and may therefore, influence the clinical outcome of HIV-1.

In summary, the studies reported here should prove useful as a basis for further studies that will focus on analyzing the differential role of the CCR5 mRNAs, if any, in protein expression, and delineating the protein factors and DNA sequences that are specifically responsible for transcription of CCR5 at physiologic sites. Additional studies will be required to determine the functional importance of the polymorphisms, if any, in the noncoding regions of CCR5 that we have described. These future studies should provide not only insights into the evolutionary rationale for such a complex system of multiple tran-
scripts directed by different regulatory regions, but may also spawn new ideas regarding mechanisms by which CCR5 gene expression can be down-regulated in disease states such as HIV-1 infection.

Acknowledgments—We thank Drs. R. A. Clark, A. Valente, and S.-L. Li for helpful discussions and critical reading of the manuscript. We thank the staff of the Bone Marrow Transplant Unit and the General Clinical Research Center (Audie Murphy VA Medical Center, San Antonio, TX) for their help in apheresis of the normal donors. We thank Nancy Nichols for help with automated DNA sequencing.

REFERENCES

1. Samson, M., Labbe, O., Mollereau, C., Vassart, G., and Parmentier, M. (1996) Biochemistry 35, 3362-3367
2. Combadiere, C., Ahuja, S. K., Tiffany, H. L., and Murphy, P. M. (1996) J. Leukocyte Biol. 60, 147–152
3. Raport, C. J., Gosling, J., Schweikart, V. L., Gray, P. W., and Chao, I. F. (1996) J. Biol. Chem. 271, 17161–17166
4. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) Science 272, 1955–1958
5. Dragic, T., Litwin, V., Choe, H., Sun, Y., Sullivan, N., Rolls, B., Ponath, P. D., Wu, L. J., Mackay, C. R., Larosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) Cell 85, 667–673
6. Deng, H. K., Liu, R., Ellmier, W., Choe, S., Unutmaz, D., Burkhart, M., Dimarzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996) Nature 381, 661–666
7. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rolls, B., Ponath, P. D., Wu, L. J., Mackay, C. R., Larosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) Cell 85, 1135–1148
8. Doranz, B. J., Rucker, J., Yi, Y. J., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Colman, R. G., and Doms, R. W. (1996) Cell 85, 1149–1158
9. Li, B., Paxton, W. A., Catravas, J., Ceradini, D., Martin, S. R., Burak, R. E., Macdonald, M. E., Stuhlman, H., Koup, R. A., and Landau, N. R. (1996) Cell 86, 367–377
10. Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, R. A., Saragost, S., Lapoumeroulle, C., Cegnas, J., Forcelli, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Sana, Y., Yi, Y. J., Routledge, M., Oro, S., Efstathiou, S., and Avvedimento, E. V. (1996) J. Immunol. 156, 4213–4222
11. Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J., Li, J., Omi, T., Kajii, E., and Koup, R. A. (1996) J. Exp. Med. 184, 1681–1691
12. Murphy, P. M. (1994) Annu. Rev. Immunol. 12, 593–633
13. Murphy, P. M. (1996) Cytokine Growth Factor Rev. 7, 47–64
14. Iwamoto, S., Li, J., Omi, T., Ikemoto, S., and Kajii, E. (1996) Blood 88, 378–385
15. Iwamoto, S., Li, J., Omi, T., Kajii, E., and Ikemoto S. (1996) Blood 88, 622–626
16. Ahuja, S. K., Shetty, A., Tiffany, H. L., and Murphy, P. M. (1994) J. Biol. Chem. 269, 26381–26389

17. Iwamoto, S., Li, J., Omi, T., Kajii, E., and Ikemoto S. (1995) Science 267, 272–276
18. Ahuja, S. S., Brown, M. R., Fleisher, T. A., Ahuja, S. K., and Malech, H. L. (1996) J. Immunol. 156, 4345–4353
19. Alkhatib, G., Ahuja, S. S., Light, D., Mummidi, S., Berger, E. A., and Ahuja, S. K. (1997) J. Biol. Chem. 272, 19771–19776
20. Prentice, D. S. (1991) Comput. Appl. Biosci. 7, 203–206
21. Quandt, K., Prech, K., Karas, H., Wingender, E., and Werner, T. (1995) Nucleic Acids Res. 23, 4878–4884
22. Orkin, H. S. (1995) J. Biol. Chem. 270, 4955–4958
23. Ayouhi, T. A. Y., and Van De Ven, W. J. M. (1996) FASEB J. 10, 453–460
24. Iwamoto, S., Li, J., Omi, T., Kajii, E., and Ikemoto S. (1996) Science 276, 272–276
25. Ball, H. J., Shine, J., and Herzog, H. (1995) J. Biol. Chem. 270, 27272–27276
26. Bibbs, R. J. B., Wyley, S. M., Pragnell, I. B., and Graham, G. J. (1997) J. Biol. Chem. 272, 12495–12504
27. Jones, R. K. (1993) Cell 74, 9–14
28. Kazak, M. (1989) J. Cell Biol. 108, 229–241
29. Kazak, M. (1991) J. Cell Biol. 115, 887–893
30. Oliveira, C. C., and McCarthy, J. E. G. (1995) J. Biol. Chem. 270, 8936–8943
31. Parola, A. L., and Kubik, B. K. (1994) J. Biol. Chem. 269, 4497–4505
32. Kazak, M. (1996) Proc. Natl. Acad. Sci. U. S. A. 83, 2850–2854
33. Smale, S. T., and Baltimore, D. (1989) Science 244, 48–52
34. Anarsi-Lari, M. A., Liu, X.-M., Metzker, M. L., Rut, A. R., and Gibbs, R. A. (1997) Nature Genet. 15, 221–222
35. Leen, M. P., Oguta, E. U., and Gorski, J. (1994) Hum. Immunol. 41, 112–120
36. Sloan, J. H., Hasegawa, S. L., and Boss, J. M. (1992) J. Immunol. 148, 2591–2599
37. Angotti, E., Mele, E., Costanzo, F., and Arvedi, D. E. (1994) J. Biol. Chem. 269, 17371–17374
38. Naganawa, S., Ginsberg, H. N., Glickman, R. M., and Ginsburg, S. G. (1997) J. Clin. Invest. 99, 1958–1965
39. Song, Z., Casolaro, V., Chen, R., Georas, S. N., Monos, D., and Ono, S. J. (1996) J. Immunol. 156, 424–429
40. Iwamoto, S., Nakajima, T., Williams, C. S., Quackenbush, J., Puryear, R., Powers, M. C., Cheng, T., Ludwig, E. H., Sharma, A. M., Hata, A., Jeunemaitre, X., and Lalouel, J. M. (1997) J. Clin. Invest. 99, 1786–1797
41. Dallinga-Thie, G. M., van Lindt-Sibenius Trip, M., Rotter, J. I., Cantor, R. M., Bu, X., Luisi, A. J., and de Bruin, T. W. (1997) J. Clin. Invest. 99, 953–961
42. Kazazian, H. H. (1990) Semin. Hematol. 27, 209–223
43. McGuire, W., Hill, A. V. S., Allsopp, C. E. M., Greenwood, B. M., and Kwaikutowski, D. (1994) Nature 371, 508–511
44. Triola, A., Dragic, T., Arthos, J., Binyo, G., Fifield, M., Olen, W. C., Allaway, G. P., Cheng-Mayer, C., Robinson, J., Maddon, P. J., and Moore, P. J. (1996) Nature 384, 184–187