SHORT COMMUNICATION

Localization of the Human CXC Chemokine Subfamily on the Long Arm of Chromosome 4 Using Radiation Hybrids

William S. Modi[1] and Zhang-Qun Chen

Intramural Research Support Program, SAIC Frederick, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

Received June 9, 1997; accepted October 17, 1997

All 12 of the human CXC chemokine genes were physically mapped using gene-specific PCR primers and the GenBridge 4 radiation hybrid panel. Nine genes, PF4, PF4V1, GRO1, GCP2, PPBP, IL8, GRO2, GRO3, and SCYB5, were assigned within a 1.8-cR interval of one another on 4q. Two additional genes, MIG and INP10, map within 0.5 cR of each another and 6 cR distal to the above-mentioned group. The final gene, SDF1, is localized on 10q. Phylogenetic analyses of amino acid sequences revealed that SDF1 is the most divergent member and that the physically separated MIG–INP10 pair constitutes a distinct evolutionary lineage.

Chemokines are small basic heparin-binding proteins that activate and stimulate the infiltration of circulating lymphocytes into tissue from peripheral blood and serve as mediators of inflammation. They are assigned to two subfamilies depending upon whether the first two cysteines are (the CXC or α subfamily) or are not (the CC or β subfamily) separated by an intervening amino acid. The CXC subfamily localizes to the long arm of chromosome 4, whereas the CC group maps to chromosome 17. The CXC members primarily mediate the migration of neutrophils, while the CC group is chemotactic for a broader array of cell types, including lymphocytes, eosinophils, basophils, and monocytes. The cDNAs for these secreted proteins encode a peptide 100–125 amino acids in length, which includes a signal peptide sequence of about 20–25 amino acids, in a single open reading frame. The mature proteins contain 4 cysteine residues that mediate two intrastrand disulfide bonds, short N-terminal regions, and long C-termini. Within each subfamily the proteins share about 25–80% amino acid similarity with one another (2). Both groups of chemokines have been implicated in various disease conditions. For example, increased levels of IL8 have been shown in several inflammatory skin diseases (13), while the GRO genes are known to be differentially expressed in rheumatoid arthritis (6). Recently RANTES, MIP-1α, and MIP-1β were demonstrated to exhibit antiviral effects by interfering with HIV infection (3).

Currently 13 human CXC proteins are known. Two of these, the SDF1 α and β sequences, differ from each other due only to alternative splicing, the β gene utilizing an extra exon, which contains four additional amino acids, that is not included in the α protein (14). As a result, DNA sequence information exists for 12 distinct genes, all of which were analyzed in the present report. Earlier studies have revealed that all of these genes, except for GCP2, have been assigned to a specific chromosome. To our knowledge, this is the first chromosomal assignment for GCP2. Among the previously localized genes, all are positioned at 4q13–q21 (summarized in 2), except for SDF1, which is located at 10q11.2 (14). The goal of the present study is to refine the physical mapping data for the human CXC chemokine subfamily.

Chromosomal mapping was carried out by using PCR-based assays with the GenBridge 4 (GB4) radiation hybrid panel, which was obtained from Research Genetics (5). PCR primers were made for all 12 genes using sequences derived from GenBank (Table 1). These PCR products represent novel STSs for each of these genes. The nucleotide sequences of each PCR product were determined to ensure that the appropriate gene region was being amplified. Polymerase chain reactions were carried out as follows. Twenty-five nanograms of hybrid or control DNA was amplified in a 10-μl volume in a reaction buffer consisting of 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each dNTP, 1 pmol of each primer, and 0.001 units of *Taq* Gold (Perkin–Elmer) polymerase. The

[1] To whom correspondence should be addressed. Telephone: (301) 846-1303. Fax: (301) 846-1909. E-mail: modi@ncifcrf.gov.

GENOMICS **47,** 136–139 (1998)
ARTICLE NO. GE975100
0888-7543/98 $25.00

TABLE 1

Localization of the CXC Chemokine Genes Using the GenBridge 4 Radiation Hybrid Panel

Gene symbol	Accession No(s).	Map position	Primer sequences	PCR product size (bp)
GCP2	U83303	485.62	GGA AAG CCT ACG CTT CTC CC TCC ATA CGA AAT CCA CAG CC	239
PF4V1	M26167	485.62	GAG GTA TGC ATC TTT GTA C GTG CGC AGG GAA AGT CGG GC	243
PF4	M25897	486.38	CCT GAA GGT CTA TCT GAG AC TGG GGA TGC CTG GTG TAA AC	221
GRO1	J03561 X54489	486.60	ATA TTT CTG AGG AGC CTG CAA C CCC TTT GTT CTA AGC CAG AAA C	300
PPBP	M54995	487.04	GTT CTG TTT CTG CCA AAC TTC AGA GGG TTG AAA CCA GGC TTA T	216
IL8	Y00787 M28130	487.04	ATG CCA CTG AAA CTT CAA GC GTT CGG ATA TTC TCT TGG CC	292
GRO2	M36820 M53799	487.15	CTT CGT GAT GAC ATA TCA CAT GTC CTC TGC TCT AAC ACA GAG GGA AAC	294
GRO3	M36821	487.26	TTA TCA GCG TAT CAT TGA CAC GTT TAA TGA CAC ATT TAA AAC	279
SCYB5	X78686	487.37	GAG CAC GCA TGG AAA AGT TTC GAC ACT TTG TAA GAT TAA CAG C	322
MIG	X72755	493.44	GGC CTC ACA CAC AAT GTG TC AAT GGG AGA AGA AAG CAC TTC	313
INP10	X02530	493.76	CTA AAA GGT GAC CAA TGA TGG TC GGT ACT AAG GAA TCT TTC TGC	302
SDF1	L36033	316.61	CAG GAC ATT TCT CTA AGA GAA C TTC CTA CTT CCT ACA TGT TAC	290

Note. Positions given are in centirays from the top of chromosome 4 (all genes except SDF1) or chromosome 10 (SDF1) radiation hybrid maps (5).

PCR cycling conditions used for all primer pairs were as follows: an initial 94°C denaturation step for 10 min followed by 35 cycles of 94°C denaturation for 30 s, 57°C annealing for 1 min, and 72°C extension for 1 min, followed by a 72°C heating for 5 min. PCR products were run out in 1.2% agarose gels and stained with ethidium bromide.

After typing of the panel of radiation hybrids for each primer pair, the resulting vectors were sent by electronic mail to the MIT/Whitehead Institute Genome Center for analysis. Results of the mapping are presented in Table 1. The SDF1 genes localize to a position 5.66 cR from marker AFMB032YH1 on the long arm of chromosome 10. This corresponds to a map location 316.61 cR from the top of the chromosome 10 radiation hybrid map (5). The remaining 11 genes all reside on the long arm of chromosome 4 and may be assigned to two primary groups based upon their location. Nine (GCP2, PF4V1, PF4, GRO1, PPBP, IL8, GRO2, GRO3, and SCYB5) of the 11 were assigned to an interval 1.8 cR in length, in a region 8 to 10 cR distal to the reference marker WI-4767. This corresponds to the approximate interval between 485 and 487 cR from the top of the chromosome 4 radiation hybrid map. The 2 remaining genes, MIG and INP10, are each located about 6.5 cR farther distally from the former group and from the reference marker WI-5565 (Table 1).

The statistical analysis of radiation hybrids estimates the probabilities of retaining certain chromosomal fragments. Unfortunately, it is difficult to calculate physical distances or confidence intervals accurately from radiation hybrid mapping data (8). As a result, it is not possible to definitively determine gene order among closely spaced markers. However, this entire group of 11 genes covers the interval from about 485 to 494 cR, and given the general relationship of 2.9 cR/Mb along chromosome 4 (7), these genes span a region that is approximately 2.75 Mb in size. The tentative order for the markers/genes under discussion is centromere–WI-4767–(GCP2, PF4V1, PF4, GRO1, PPBP, WI-5565, IL8, GRO2, GRO3, SCYB5)–(INP10–MIG)–telomere.

Several of these genes have been previously mapped using various procedures. Pulsed-field gel electrophoresis colocalized PPBP, PF4, GRO1, and IL8, but not INP10, to a 700-kb *Sfi*I fragment, and further, PPBP, PF4, and GRO1 were assigned to a 400-kb *Sac*II band. Analyses of genomic clones suggested that PPBP is located only 7 kb upstream of PF4. Gene-specific probes for GRO2, GRO3, SCYB5, or MIG were not examined (15). Another report determined that MIG and INP10 were located only 16 kb apart (9). GRO2 (Accession No. X53799), also known as MIP-2α, was assigned to position 484.62 cR on the GB4 panel and further placed on the singly linked YAC contig WC4.2 as the STS reference marker WI-9034 (5). A GB4 vector for PF4 named RH17684, submitted by Généthon, localizes this gene at position 484.58 cR (7.58 cR from WI-4767). A

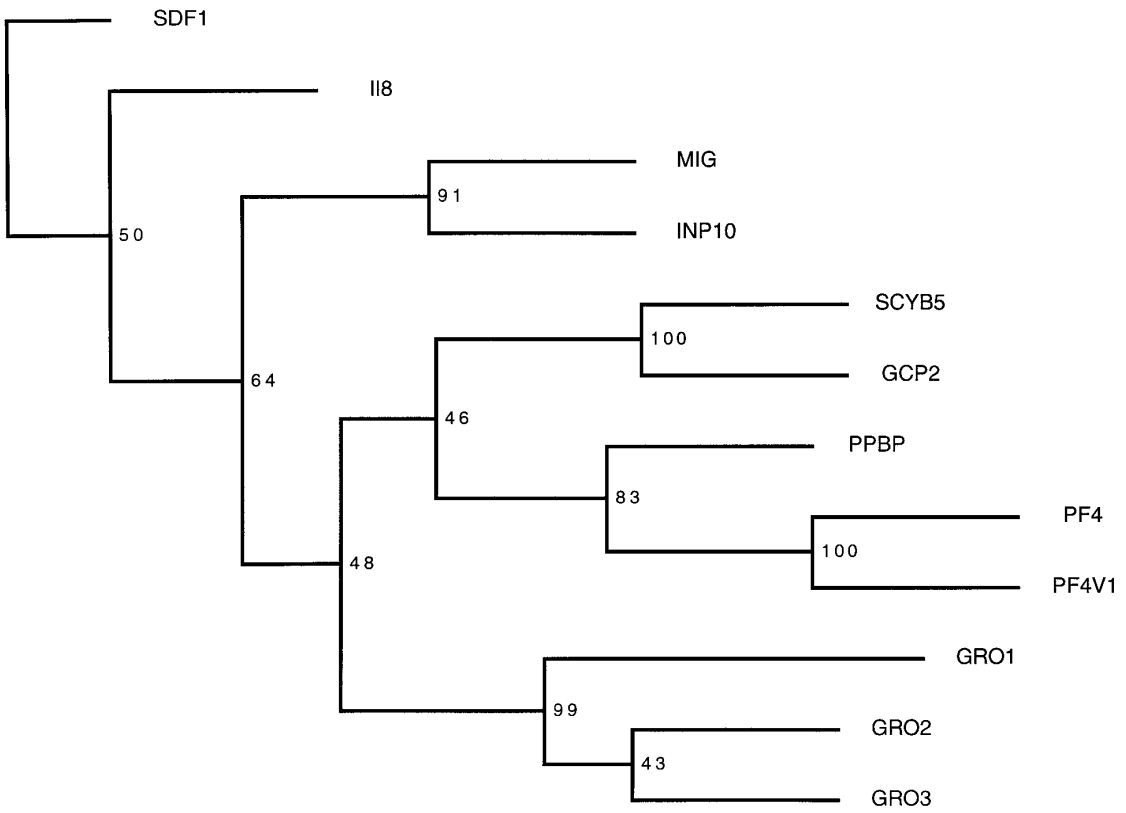

FIG. 1. Neighbor-joining phylogenetic tree of the human CXC chemokines based upon 66–70 amino acids from the mature protein region constructed using Dayhoff Pam distances and the NEIGHBOR algorithm from the computer package PHYLIP (4). Values at nodes represent bootstrap support following 1000 replications. A more extensive discussion of the phylogenetic analysis is available (W. S. Modi and Yoshimura, unpublished).

GB4 vector prepared by the Sanger Centre, called RH12888, assigns the IL8 gene at 483.83 cR (6.83 cR from WI-4767). An STS for the PF4V1 gene, called WI-7208 or D12S2028, has incorrectly been assigned to chromosome 12 (5). An EST for the SDF1 gene, EST1552, localizes to position 316.28 cR on chromosome 10, and an STS WI-9393 has been assigned to the singly linked YAC contig WC10.2 (5). The positions determined from the present study differ from some of those in the literature by as much as 3 cR for the same gene. Two explanations may account for these discrepancies: (1) The Whitehead Institute database is continually updated by the addition of new markers. Since marker density affects the probability of assigning an unmapped locus, an identical locus mapped at two different points in time may result in slightly different positions and (2) variation in the polymerase chain reactions may lead to faint bands that are arbitrarily scored as present or absent by different investigators or under different experimental conditions.

This is the first study in which all known human CXC chemokine genes have been physically mapped using a common procedure. Nine of the genes, of which the order relative to one another cannot be determined with any confidence, are located within about a 1.8-cR interval. The 2 remaining genes, INP10 and MIG, are positioned distally to this cluster. An evolutionary tree diagram based upon the amino acid sequences for these proteins is presented in Fig. 1. With this information, we may address the question of whether the proteins that share the greatest amino acid similarity are located closer together along the chromosome. The phylogenetic tree reveals that SDF1 and IL8 are the most divergent sequences in the group, that the three GRO proteins share a high degree of amino acid similarity, and that the remaining six sequences fall into three groups, one containing PPBP, PF4, and PF4V1, another containing SCYB5 and GCP2, and the final one comprising MIG and INP10. The most divergent protein of all, SDF1, is located on a different chromosome. The second most divergent sequence, IL8, is physically positioned closely to 8 other family members and does not support the idea that the more divergent sequences would be more distantly localized than more closely related sequences. However, two of the proteins that form a distinct lineage on the tree, MIG and INP10, are physically separated from the other 9 genes. Therefore, with the exception of IL8, the more similar sequences appear to be closer together.

Given the similarities in intron/exon structure, amino acid sequence, and immunological function, it

has been suggested that these chemokine genes are all derived through tandem gene duplication from a common ancestral gene that was presumedly located on chromosome 4. The position of SDF1 on chromosome 10 must then represent a translocation event. Interestingly, 10 of the CC chemokines have also been shown to be clustered very closely together on the long arm of chromosome 17. These genes fall into a contig comprising three YAC clones that cover an estimated distance of several megabases (10). However, one CC chemokine gene, TARC, was recently placed on the map at 16q13 (11). This gene was apparently translocated from chromosome 17 to 16. Similarly, the receptors for the CC chemokines are localized within an estimated 8.5 Mb of one another at 3p21.3–p24 (12), and the IL8 α and β receptor genes have been mapped within 150 kb of each other on the long arm of chromosome 2 (1). These findings reveal the role of tandem gene duplication in the evolution of these functionally and evolutionarily related multigene families. Additional efforts will reveal whether or not any of the genes within any of these clusters are coordinately regulated.

REFERENCES

1. Alvarez, V., Coto, E., Seiten, F., and Lopez-Larrea, C. (1994). A physical map of two clusters containing the genes for six proinflammatory receptors. *Immunogenetics* **40:** 100–103.
2. Baggiolini, M., Dewald, B., and Moser, B. (1994). Interleukin-8 and related chemotactic cytokines—CXC and CC chemokines. *Adv. Immunol.* **55:** 97–179.
3. Cocci, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995). Identification of RANTES, MIP-1a, and MIP-1b as the major HIV-suppressive factors produced by $CD8^+$ T cells. *Science* **270:** 1811–1815.
4. Felsenstein, J. (1993). PHYLIP (phylogeny inference package) 3.5c manual. Univ. of Washington, Seattle.
5. Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillet, D., Muselet, D., Prud'Homme, J-F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J., and Goodfellow, P. N. (1996). A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5:** 339–346.
6. Hogan, M. B., Sherry, C., Ritchlin, M., Fabre, R., Winchester, A., Cerami, A., and Bucala, R. (1994). Differential expression of the small inducible cytokines Groα and Groβ by synovial fibroblasts in chronic arthritis: Possible role in growth regulation. *Cytokine* **6:** 61–69.
7. Hudson, T. J., *et al.* (1995). An STS-based map of the human genome. *Science* **270:** 1945–1954.
8. Jones, H. B. (1997). Estimating physical distances from radiation hybrid mapping data. *Genomics* **43:** 258–266.
9. Lee, H-H., and Farber, J. M. (1996). Localization of the gene for the human MIG cytokine on chromosome 4q21 adjacent to INP10 reveals a chemokine 'mini-cluster.' *Cytogenet. Cell Genet.* **74:** 255–258.
10. Naruse, K., *et al.* (1996). A YAC contig of the human CC chemokine genes clustered on chromosome 17q11.2. *Genomics* **34:** 236–240.
11. Nomiyama, H., Imai, T., Kusuda, J., Miura, R., Callen, D. F., and Yoshie, O. (1997). Assignment of the human CC chemokine gene TARC (SCYA17) to chromosome 16q13. *Genomics* **40:** 211–213.
12. Samson, M., Soularue, P., Vassart, G., and Palmentier, M. (1996). The genes encoding the human CC-chemokine receptors CC-CKR1 to CC-CKR5 (CMKBR1–CMKBR5) are clustered in the p21.3–p24 region of chromosome 3. *Genomics* **36:** 522–526.
13. Schroder, J-M. (1992). Generation of NAP-1 and related peptides in psoriasis and other inflammatory skin diseases. *In* "Interleukin-8 (NAP-1) and Related Chemotactic Cytokines" (C. Sorg, Ed.), pp. 54–76, Karger, Basel.
14. Shirozu, M., Nakano, T., Inazawa, J., Tashiro, K., Tada, H., Shinohara, T., and Honjo, T. (1995). Structure and chromosomal localization of the human stromal cell derived factor 1(SDF1) gene. *Genomics* **28:** 495–500.
15. Tunnacliffe, A. , Mujumdar, S., Yan, B., and Poncz, M. (1992). Genes for β-thromboglobulin and platelet factor 4 are closely linked and form part of a cluster of related genes on chromosome 4. *Blood* **79:** 2896–2900.