Structural analysis of the mouse chromosomal gene encoding interleukin 4 which expresses B cell, T cell and mast cell stimulating activities

Takeshi Otsuka, Douglas Villaret, Takashi Yokota, Yutaka Takebe, Frank Lee, Naoko Arai and Ken-ichi Arai*

Department of Molecular Biology, DNAX Research Institute of Molecular and Cellular Biology, 901 California Ave., Palo Alto, CA 94304, USA

Received August 19, 1986; Revised and Accepted November 21, 1986

ABSTRACT
Based on homology with the mouse interleukin 4 (IL-4) cDNA that expresses B cell, T cell, and mast cell stimulating activities (Lee, F. et al., (1986) Proc. Natl. Acad. Sci. USA 83, 2061), we have isolated from a Balb/c mouse liver DNA library the mouse chromosomal gene and analyzed its overall structure. The gene occurs as a single copy in the haploid genome and contains four exons and three introns. The exon sequences almost completely match the cloned cDNA sequence. Interestingly, there is a fairly high degree of homology between mouse IL-4 and mouse IL-2 genes extending more than 200 bp upstream of a "TATA" like sequence located 20 bp upstream of the transcription initiation site. These sequences may play an important role in the regulated expression of this gene in concanavalin A or antigen-stimulated T lymphocytes. The supernatant of COS7 cells transfected with plasmid DNA containing the entire gene exhibited both T cell growth factor and mast cell growth factor activities.

INTRODUCTION
We have previously reported the isolation of a cDNA encoding an interleukin with B cell, T cell and mast cell stimulating activities from a cDNA library constructed with mRNA from concanavalin A (ConA) stimulated mouse T cell clone Ly1+2-/9 (1). The T cell growth factor (TCGF) and mast cell growth factor (MCGF) activities of this clone were distinct from either the TCGF activity of IL-2 or the MCGF activity of IL-3. The B cell stimulating activities of this lymphokine include co-stimulation of anti-IgM-activated B cells (2), induction of Ia antigen on resting B cells (2) and enhancement of IgE and IgG1 production (3), all properties known to be associated with B cell stimulatory factor 1 (BSF-1) (4-8). The same cDNA clone has been isolated by others based on IgG1 enhancing activity (9). Because of the multiple biological activities, we suggested that
this lymphokine be called interleukin 4 (IL-4) and we will use this term in this paper. It should be noted, however, that the same name has been proposed for another lymphokine apparently possessing characteristics different from this factor (10).

Although IL-4 possesses some biological activities similar to IL-2 and IL-3, there are no significant homologies between IL-4 and either IL-2 or IL-3 cDNAs in nucleotide sequence or in predicted amino acid sequences. Mouse helper T cell clones have been classified into two subsets based on their lymphokine production upon activation (11). IL-4 is produced by the TypeII (T\text{H}2) subset which also produces GM-CSF and IL-3 but not IL-2 and IFN-\gamma. On the other hand, TypeI (T\text{H}1) cells produce IL-2, IFN-\gamma, GM-CSF and IL-3 but not IL-4 (11). Thus, these lymphokines seem to be produced under precise control mechanisms in each T cell subset. Although the regulation of their expression may be at the transcriptional level (T cell lymphokine transcripts were not detectable without stimulation by lectin or by antigen), little is known about the mechanism of the regulated expression of these lymphokine genes.

In this paper, we report the isolation and the structural and functional analyses of the mouse IL-4 gene.

MATERIALS AND METHODS

Southern blotting of mouse chromosomal DNA.

Mouse chromosomal DNA was isolated from Balb/c mouse liver, Balb/c mouse thymus and the mouse helper T cell clone Ly1^+2^-/\theta, established from a C57BL/6 mouse. Each high molecular weight DNA sample was digested with BamHI or EcoRI, electrophoresed on a 0.6% agarose gel and transferred to a nitrocellulose filter according to the method of Southern (12). The RsaI fragment (373 bp) of mouse IL-4 cDNA clone 2AE3 (1) was labeled with \textsuperscript{32}P by nick translation and used as a probe. Hybridization was in 6x SSPE containing 50% formamide at 42°C; the filter was washed in 0.2x SSPE at 65°C.

Cloning and sequencing of the mouse IL-4 gene.

A Balb/c mouse liver genomic DNA library, constructed in bacteriophage \lambda Charon 4A, was kindly provided by Dr. Mark Davis. Phage plaques of this library were screened with \textsuperscript{32}P-labeled
mouse IL-4 cDNA as a probe. Two independent phage clones were isolated and appropriate fragments were subcloned into the plasmid vector pUC18. After estimating the location of coding regions by Southern hybridization, appropriate fragments were subcloned into pUC18 or pUC19 and were directly subjected to sequence analysis by the dideoxy chain termination method with supercoiled DNA templates (13).

Nuclease S1 mapping.

The 5' end-labeled TaqI/HindIII fragment which contains the 5' flanking region and exon 1 of the mouse IL-4 gene was cut with PstI and the TaqI/PstI fragment (positions -298 ~ 107) was isolated from polyacrylamide gel. This 5' end-labeled DNA (10,000 c.p.m.) was mixed with 1 μg of polyA+ RNA from a cloned helper T cell, clone Ly1+2~/9 stimulated for 9 hr with ConA and incubated at 75°C in hybridization buffer for 10 min. Hybridization was then performed at 48°C for 16 hr. This mixture, after adding nuclease S1 in prechilled buffer, was incubated at 37°C for 30 min (14).

The 5' end-labeled TaqI/PstI fragment described above, which encodes the antisense sequence of the mouse IL-4 gene, was also subjected to chemical cleavage using the methods of Maxam and Gilbert (15).

The sample treated with nuclease S1 and the sequencing samples were subjected to electrophoresis on a 6% polyacrylamide-urea gel.

Transfection in COS7 cells and bioassays for TCGF and MCGF activities.

1x10^6 COS7 cells in 10 ml of Dulbecco's Modified Eagle Medium were transfected with 20 μg of plasmid DNA by the DEAE dextran method (16). The supernatant was harvested after 3 days and assayed for TCGF activity and for MCGF activity by using the HT-2 T cell line and the MC/9 mast cell line, respectively (17). Proliferation was determined using a colorimetric assay (18).

RESULTS

Southern blotting analysis of chromosomal DNA.

To elucidate both the structure of the mouse IL-4 gene and the number of genes in the mouse genome, Southern blotting
Fig. 1 Southern blotting of mouse chromosomal DNA with mouse IL-4 cDNA probe. High molecular weight DNA was prepared from Balb/c mouse liver, Balb/c mouse thymus and helper T cell clone Ly1+2-9. Each DNA sample was digested with either BamHI (lane A) or EcoRI (lane B). Hybridization was done by using ^32P-labeled Rsal fragment of mouse IL-4 cDNA (1x10^8 c.p.m./μg) as a probe under conditions described in MATERIALS AND METHODS. Positions of size markers (kb) are shown on the left.

analysis was performed. Chromosomal DNAs isolated from Balb/c mouse liver, Balb/c mouse thymus and helper T cell clone Ly1+2-9 were examined. As shown in Fig. 1, chromosomal DNA from three different sources showed the same hybridization pattern, indicating that the IL-4 gene contains a single EcoRI site but no BamHI site. The results also show that there is only a single copy of the IL-4 gene per mouse haploid genome and that no detectable rearrangement has occurred in the helper T cell clone Ly1+2-9 which expresses IL-4 after ConA stimulation.
Fig. 2 Restriction cleavage map of the mouse IL-4 gene region. λMIL9-3 and λMIL16-1 were isolated from the λ charon 4A mouse liver library. A 5.5 kb HindIII fragment and a 7.5 kb EcoRI fragment were subcloned into pUC18 (pMIL5 and pMIL1); both subclones were used to make this cleavage map. A 9 kb region which includes the entire mouse IL-4 gene is expanded. Arrows, direction and extent of sequence analysis; black boxes at the bottom, four exons; B, BamHI; B2, BglII; E, EcoRI; H, HindIII; P, PstI.

Isolation of IL-4 gene from genomic library and its overall structure.

A Balb/c mouse liver genomic DNA library was screened by using 32P labeled mouse IL-4 cDNA clone 2AE3. Two independent phage plaques, λMIL9-3 and λMIL16-1, were isolated and the insert DNA from both clones was used to determine a restriction endonuclease cleavage map (Fig. 2). A 7.5 kb EcoRI fragment and a 5.5 kb HindIII fragment were isolated from λMIL9-3 and subcloned into pUC18 to yield pMIL1 and pMIL5, respectively. Subsequently, Southern blotting was performed against these two subclones following digestion with several restriction endonucleases to identify the location of coding regions. Based on these data, we subcloned appropriate fragments into pUC18 and sequenced these by the dideoxy chain termination method using alkaline denatured supercoiled plasmid DNA according to the strategy shown in Fig. 2; the sequence is presented in Fig. 3. Transcription initiation site.

To determine the transcription initiation site, S1 nuclease
Nucleic Acids Research

Fig. 3 Nucleotide sequence of mouse IL-4 gene. Four exons are framed by the thick line. Positive numbers refer to nucleotide positions of exons and negative numbers show nucleotide positions of 5' flanking region from the transcription initiation site (*). "TATA" box (under line -27 to -20).

mapping was performed using the 5' end-labeled TaqI/PstI fragment and Lyt1+2-/+ T cell polyA+ RNA. RNA-DNA hybrids were electrophoresed alongside the sequencing ladder of the same
Fig. 4 Nuclease S1 mapping of mouse IL-4 mRNA in activated helper T cell clone Ly1\(^+\)/9. The products of nuclease S1 treatment against DNA/RNA hybrid of 5' end labeled TaqI/PstI (-298~107) fragment and the clone Ly1\(^+\)/9 mRNAs were subjected to electrophoresis together with samples for DNA sequence analysis of the same TaqI/PstI fragment by Maxam-Gilbert's method (complementary to the IL-4 coding sequence) on a 6% polyacrylamide-urea gel. The nucleotide pairs at the right hand side represent the corresponding sequence of the mouse IL-4 gene.

Taking into consideration the 1.5 base difference of migration between the S1 protected fragment and the corresponding chemical cleavage fragments, we estimate the transcription initiation site to be at the position marked with an asterisk in Fig. 3.

Nucleotide sequence of mouse IL-4 gene.

The nucleotide sequence of the gene was determined for each exon and its surrounding regions and the 5' flanking region. This gene extends about 6 kb and is composed of four exons and three introns. Based on the location of the transcription initiation site determined by nuclease S1 mapping, the length of exon 1 is 195 bp. Exons 2, 3 and 4 are 48 bp, 153 bp and 191 bp, respectively. Each intron interrupts the reading frame precisely between codons and begins with the consensus sequence "GT" and ends with the consensus sequence "AG". A "TATA" like sequence,
Fig. 5 Structure of an expression vector for IL-4 gene. A 4.7 kb BamHI/EcoRI fragment and a 7.5 kb EcoRI fragment from λMIL9-3 were ligated and inserted between BglII and EcoRI sites of plasmid pB324, which is made by the converting the XhoI site of plasmid pL1 to a BglII site. This recombinant plasmid DNA was designated pGIL1. The SV40 early promoter region is shown by the hatched box. Dark boxes and open boxes indicate exons and introns of the IL-4 gene, respectively.

TATATATA, is located 20 bp upstream of the transcription initiation site. The sequence CCAATGT at positions -70 ~ -64 is very similar to "TCCATGT" which is located 103 bp upstream of the putative cap site of the mouse IL-3 gene and also is similar to another consensus sequence "GGCCAATCT" (19).

The sequence of the coding region differs from the previous cDNA sequence data at two positions. The sequence ACGGCA (positions 47~52) was assigned ACGCA in error. The other discrepancy is a single base difference of C instead of T at position 291 representing a genuine difference, which does not, however, alter the amino acid sequence. Since IL-4 cDNA clone 2AE3 was derived from C57BL/6 mice and the genomic DNA was from Balb/c mice, this sequence discrepancy may be due to strain differences, as was found in the case of the mouse GM-CSF gene sequence (20).

Expression of TCGF and MCGF activities by IL-4 gene.

In order to examine whether the mouse IL-4 gene we have isolated is functional, a 4.7 kb BamHI/EcoRI fragment and a 7.5 kb EcoRI fragment were ligated and inserted between the BglII and EcoRI sites of plasmid pB324, which is made by converting XhoI site to BglII in the mammalian expression vector plasmid pL1 (Fig. 5) (21). This reconstructed plasmid DNA, pGIL1, contains
Fig. 6 Biological activities of the supernatant from COS7 cells. (A) TCGF activity was determined with HT-2 cells by using the colorimetric assay. (B) MCGF activity was determined with MC/9 cells by using the colorimetric assay. Samples: 1, 2A-E3 COS7 supernatant (O---O). 2, pGIL1 COS7 supernatant (●---●). 3, mock-transfected COS7 supernatant (Δ---Δ).

The data presented in this paper show that a single copy of the IL-4 gene exists in a haploid genome as is the case for other lymphokine genes such as IL-2 (22,23), IL-3 (24), GM-CSF (25), and IFN-γ (26) which are expressed inducibly in activated helper T cells. Chromosomal DNA from different tissue sources shows the
same hybridization pattern indicating that no apparent
rearrangement of this gene takes place during the development of
T cells. The mouse IL-4 gene is composed of four exons and three
introns extending about 6 kb in length. In view of the
relatively small size of the IL-4 cDNA, the size of this gene is
relatively longer than those of other lymphokines described
above. Although the IL-4 gene has almost no obvious nucleotide
sequence homology with these lymphokine cDNAs which have been
isolated, it still shares several features with other lymphokine
genes, 1) the number of exons is four or five, 2) the second
exon, which is the shortest, is separated from the first exon by
100~250 bp, and 3) each exon-intron junction exists precisely
between codons.

Transfection of plasmid DNA carrying the mouse IL-4 gene
downstream of the SV40 early region promoter into COS7 cells
shows that the gene is functional. At present, we do not know
whether the production of IL-4 is directed by its own promoter or
by the SV40 early region promoter or both. The lower expression
level directed by the transfected IL-4 gene compared to the IL-4
cDNA might be due to the existence of a long 5' flanking region
(1 kb) between the SV40 early region promoter and the mouse IL-4
coding region, the difference of promoter strength between SV40
ey early region and mouse IL-4, copy number differences of
transfected DNA in COS7 cells or a combination of these.
An interesting feature of this gene exists in the 5' flanking region, generally thought to have a regulatory role for inducible expression of the gene. Numerous patches of homologous sequences between IL-4 and IL-2 (Fig. 7) were found in the region extending more than 200 bp upstream of the "TATA" like box, while neither mouse IL-4 nor mouse IL-2 show any convincing sequence homology with other lymphokine genes (IL-3, GM-CSF and IFN-γ) in this region. An especially high degree of homology is found at positions -1 ~ -18, positions -29 ~ -40 and positions -52 ~ -69 (Fig. 7). Among mouse lymphokine genes, IL-3 and GM-CSF genes also share homologous sequences in this region (25). Such sequences may be involved in the coordinated expression of both genes in every T helper subset after stimulation with lectins or antigen. Despite the homology of this region shared between mouse IL-4 and IL-2 genes, they are apparently expressed predominantly in different T cell subsets. Other elements may specifically control the expression of either IL-4 or IL-2.

Sequence homologies of lymphokine genes between different species are found frequently in the 5' flanking regions. For example, the 5' flanking regions of mouse and human GM-CSF genes show nearly 90% homology (25) and those of mouse IL-2 and human IL-2 show 85% homology (23). In spite of the divergence between mouse and rat IL-3 coding sequences, the 5' flanking regions still show extensive sequence homology (27). From this viewpoint, it is tempting to speculate that the sequence homology in the 5' flanking regions between IL-2 and IL-4 genes may indicate their evolutionary relationship.

ACKNOWLEDGEMENT

We thank M. Davis for providing the mouse liver DNA library and H. Hagiwara for advice on nuclease S1 mapping analysis.

*To whom correspondence should be addressed

REFERENCES

2. Roehm, N. W., Leibson, H. J., Marrack, P., Cambier, J. C.,


