Organization and structure of the mouse interleukin-2 gene

Akira Fuse*, Takashi Fujita, Hidetaro Yasumitsu, Nobukazu Kashima+, Katsushige Hasegawa and Tadatsugu Taniguchi[§]

Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Toshimaku, Tokyo 170 and Institute for Molecular and Cellular Biology, Osaka University, Suita-shi, Osaka 565, Japan

Received 10 October 1984; Revised and Accepted 20 November 1984

ABSTRACT

We have cloned a chromosomal DNA segment which covers the entire sequence for the murine interleukin-2 gene and analysed the structure of the gene. The coding regions are separated into four blocks by three introns each of which is located similarly to the corresponding human gene. The exon sequences can be aligned perfectly with the previously cloned cDNA sequence. Of particular interests is the presence of sequences within the 5'flanking region which are highly conserved between mouse and man. The conserved region which spans more than 400 base pairs may play a role in the regulation of IL-2 gene expression.

INTRODUCTION

Interleukin-2 (IL-2) is a lymphokine produced by T cells upon antigenic or mitogenic stimulation and is required for the proliferation of T cells (1,2). Several other biological activities of IL-2 which appear to be crucial in the immune regulation have also been reported (3, 4. 5. 6. 7.) . We previously reported isolation and sequence analysis of the cDNA for human IL-2 (8), as well as the chromosomal gene (9). More recently, we have isolated a cDNA which encodes murine IL-2 (Kashima et al., submitted for publication). The cDNA contains a unique tandem repeat of CAG sequence which would encode 12 consecutive glutamine residues in the active IL-2 molecule.

In order to study the structure of the murine IL-2 chromosomal gene and its controlling region, we isolated and analysed a λ phage clone containing the gene and its flanking sequences.

MATERIALS AND METHODS

Southern blotting of total mouse DNA

Mouse chromosomal DNA was extracted from liver of BALB/c6

mouse as described before (10). High molecular genomic DNA was digested with various restriction enzymes and electrophoresed on 0.8% agarose gel. Blotting analysis of DNA was carried out by the method of Southern (11). Hybridization was carried out as described previously and filters were washed either in 3 x SSC at 65° C (lower stringent condition) or in 0.1 x SSC at 65° C (higher stringent condition).

Screening of genomic DNA library

A bacteriophage λ Charon 4A/mouse genomic DNA library prepared with partial EcoRI digests of mouse DNA from MPC 11 plasmacytoma cells was kindly provided by Dr. T. Honjo. Mouse IL-2-specific clones were screened by the method of Benton and Davis (12), using 700 bp PstI-AccI fragment of a cDNA clone, pMIL2-45 as the probe (Kashima et al., submitted for publication). Hybridization was performed as described previously(13). Positive clones were rescreened at least twice. Subcloning and sequencing of the mouse IL-2 gene

Two EcoRI fragments of 3.3 Kbp and 2.8 Kbp from the positive recombinant λ phage were subcloned into EcoRI site of plasmid pBR322. DNA segments derived from subcloned 3.3 Kbp and 2.8 Kbp fragments were labelled at either 3' end or 5' end, and subjected to sequence analysis by the chemical degradation method (14). The 0.8 Kbp EcoRI fragment from the same λ phage clone was directly subjected to sequence analysis by the dideoxy chain termination method (15).

RESULTS

Total DNA blotting analysis

In order to study structural organization of the mouse IL-2 gene, we first subjected total mouse DNA to the blotting analysis by using various probes specific for IL-2 gene. When mouse DNA was digested with various restriction endonucleases and then probed with a 7 kb human chromosomal DNA segment which contains the human IL-2 gene and its flanking region (Fig. 1 lane 1-8, ref. 9.), single positive band appeared at lower but not at higher stringent condition for washing the filters (see Figure legend). Additional bands corresponding to those observed by using mouse IL-2 cDNA probes (lane 13-17) also appeared by longer

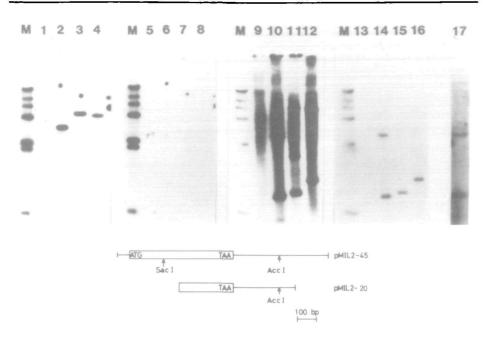


Fig. 1. Blot hybridization analysis of mouse chromosomal DNA. High molecular DNA prepared from Liver BALB/C6 mouse was digested with various restriction endonucleases (BamHI for lanes 1, 5, 9, 13 ; EcoRI for lanes 2, 6, 10, 14, 17 ; HindIII for lanes 3, 7, 11, 15 ; XbaI for lanes 4, 8, 12, 16). The resulting digests were fractionated on 0.8 % agarose gel and transferred to a nitrocellulose filter. Filters were hybridized by the published procedure (13) either with the nick-translated chromosomal DNA containing human IL-2 gene and its flanking region (total length, 7.0 kb, ref. 9) (lane 1-8) or with the nick-translated cDNA for mouse IL-2 (see figure). Filters were then washed either in 3 x SSC at 65 °C (lower stringent condition) (lane 1-4, 9-12) or 0.1 x SSC at 65 ^OC (higher stringent condition) (lane 5-8, 13-17). Lane M each contains 7 size markers with their size being 23.7 9.5 kb, 6.7 kb, 4.3 kb, 2.3 kb, 2.0 kb and 0.6 kb, kb, Brief restriction endonuclease cleavage map for respectively. the mouse IL-2 cDNAs is presented in the lower part of the figure.

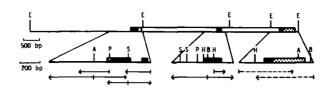
exposure of the film (data not shown). Those results suggest the presence of highly conserved sequences between human and mouse DNA either in the flanking regions or in the introns of the IL-2 gene, since the coding regions apparently show lower degree of sequence homology as evidenced in this series of blotting analysis (see below). When the PstI insert of a mouse IL-2 cDNA clone, pMIL2-20, was used as the probe, a simple pattern was obtained at higher stringent condition (lane 13-17). While the EcoRI-digested DNA gave rise to two positive bands (2.8 kb and 0.8 kb)(Fig. 1, lane 14) by this analysis, one additional band of 3.3 kb also appeared when the same DNA was probed with a longer cDNA insert from another clone, pMIL2-45 (Fig. 1, lane 17). The 3.3 kb band was similar in its size with the positive band which became detectable by probing the same DNA with the 7.0 kb human DNA probe (Fig. 1, lane 2). Since this band appeared with the cDNA probe extending further upstream, it is likely that the 5' region of the gene is located within this DNA segment (see below). Indeed, this 3.3 kb band did not appear even after longer exposure of lane 14 (result not shown). BamHI digest of the mouse DNA (Fig. 1, lane 1, 13) constantly gave a very faint signal which would correspond to a DNA larger than 15 kb. Taken together, the results suggested the presence of a single copy gene for murine IL-2. On the other hand appearance of the multiple positive bands at lower stringent washing condition (lane 9-12) indicates the presence of IL-2 related sequences within the mouse genome.

Screening of recombinant phage libraries

We next screened a gene library from partial EcoRIdigested DNA from MPC 11 cells and by using 0.8 Kbp SacI-AccI cDNA fragment as the probe and isolated 14 positive clones containing sequences specific to the mouse IL-2 gene. Three of the clones analysed all contained three EcoRI fragments whose size is in agreement with the result of blotting analysis of the chromosomal DNA as shown in Fig. 1 (lane 17). One of these is designated MIL-2G70.

Nucleotide sequence analysis

Two DNA fragments of 3.3 Kbp and 2.8 Kbp were excised from the phage clone MIL-2G70 by EcoRI digestion and they were subcloned into pBR322. DNA sequences were determined for selected regions of both inserts and compared to the known cDNA sequences (Kashima et al., submitted for publication). The strategy used for sequence analysis of the genomic DNA is presented in Fig. 2. Comparison of the mouse genomic IL-2 sequence with mouse IL-2 cDNA sequence revealed that, like the human gene, the gene is divided into four exons. A putative



Restriction map and sequencing strategy of mouse IL-2 Fig. 2. gene. Horizontal lines indicate the length of mouse DNA inserted into the λ phage Charon 4A or plasmid subclones. Filled blocks, dashed blocks and open blocks indicate protein coding regions, untranslated regions and introns, respectively. Horizontal indicate the direction and extent of arrows sequence determination without ambiguity. Dashed arrows: determination was done by the chain termination method (15) after subcloning 0.8 kb fragment into M13. Rest of the sequence the determination was carried out by the method of Maxam and Gilbert (14). A AccI site, B; BamHI site, E; EcoRI site, H; HindIII site, P; PstI site, S; SacI site.

capping site or the transcription initiation site was located 32 bp downstream from a TATAAA consensus promotor sequence (Fig. 3). The first ATG triplet was located 79 bp downstream from the TATA box. As seen also in the murine IL-2 cDNA, there is an unusual repeat of CAG triplet coding for 12 glutamine residues in a row in the first exon. The second exon (60 bp) is separated from the first exon by a short intron consisting of 97 bp. The second, the third and the fourth exons are interrupted by longer introns whose size is about 2.3 Kbp and 1.6 Kbp, respectively. As far as the available sequence data are concerned, it seems that, despite their identical location, intron sequences are distinctly dissimilar except for the junction regions between the human and There are two potential poly (A) addition mouse IL-2 genes. signals within the mouse gene (nucleotide positions 793 - 798 and 924 - 929 in Fig. 3) and, based on our sequence data for various cDNA clones, both signals seem to function and give rise to heterogeneous termini of the mRNA in the LBRM-33 cells (16).

We have also determined the sequence of about 500 bp of 5'-flanking region of mouse IL-2 gene, since (i) promoter/regulatory sequences are located in this region in many other genes of eukaryotes and (ii) this region appeared to contain sequences which show strongest cross-hybridization between human and mouse DNA around the IL-2 gene (Fig. 1). Comparison of the nucleotide sequences for the 5' flanking region

						TAC	CACC		~~ 1 5		-450	CAN	LCC N		~~ ~~~		····			лтст/	
GATTTA	-400 TICTI	TTCA	тста	TCTC	стст								-35)							
			-300												-25	0					
CATGAG					-20	0											-	150			
TEGECT							- 10	ρo												-50	
луусуу	AGGTA	ATAC	rric:	rece	ACACI	AGGT	NGAC	ren-	TGA		ATGTO	FTAA	IA10		CAT		CAC		TAT	ATT	<u>-1</u>
CAGCAT	талса	gtit	ÅÅÅŤ	ICCC.	rccc.	ATGC:	IGAAI	GAGC	TGCC	цітс.	ACCC:	FICC	TAAT	CACTO	CTC		rgaci	CTCA	GTC	TGC	GCC
ATG TA Met Ty	C AGC	ATG	CAG	стс	GCA	TCC	TCT	GTC	ACA	TTG	ACA	CTT	GTG	стс	СТТ	100 GTC	AAC	AGC	GCA	ccc	ACT
Here I y	1 941	noc	011	Lieu	A.1.6	941	Cy.	vai	1111	15		Deg	Var	Dou	Deu	vai	n #//	201	A10		
TCA AG Ser Se			Ser	Ser																	
сас ст	G GAG	CAG	20 CTG	TTG	ATG	GAC	ста	CAG	GAG	CTC	CTG	AGC	AGG	ATG	GAG	GT	AGT	GCAC	GCC.	TCC	CATC
His Le	u Glu	Gln	Leu	Leu	Het	Yab	Leu	Gln	Glu	Leu	Leu	Ser	Arg	Met	Glu	J					
TATAGG	слата	CCTT	TAGC	TTC	TGC	CAAA	GGCTN	STGT	TTAA	TAAC	CTTT	AATA	ATAA'	IGTT	ACGCT	TTC	IC AG			AGG Arg	
250						-]						
CTG AA Leu Ly															GTG	GTG	GTT	TCTG	FTTA.	ACTG	STGC
TCTAAT	G																				
									20	AA 1.	_										
									~201	00 Ы	p										
•								N	ATGT	GAAC	CTTG	TAGT	FICT	TGT	AGAT.	GGN	CVY	TAGT	CTGA	ACTTO	GTG
TATGCA	TTGGT	AGAG.		ChG		TTA	-			A A G C	- AGA		CTC M	~ T AG	AGTG	eren.		CTTA.	AGGT	-ACC	
TGGTTG	TGAGC	TCTT	CCTC	TTCT	AGAT	TAT	GGC A	ICGA	TTAC	CTCA	STCC	CCT	TTAC.	AGAG	GACA	GGN	STGG	TAAA	AGCT.	ATGTO	SCT
<u>CCTTCT</u> 300		TTAG.	AGAG	ACTG	CAGA	CTAN	CTTT	TGG	стст	TCAG	TATC	IGGT	GAGC	rgac	CTGA	GGT		CTTA	TAC	ICCTO	TA
ccc Åc.	A GAA	TTG	***	GAT	CTT	CAG	TOC		~												
									644	GAT	GAA	CIT	GGA	CCT	CTG	CGĞ	CAT	GTT	CTG	GAT	TTO
n	r Glu									Asp											
		Leu	Lys	Хsр	Leu	Gln	Сув	Leu	Glu	Asp 400	Glu	Leu	Gly	Pro	Leu	Arg	His	Val	Leu	Asp	Lei
ACT CA	A AGC	Lou	Lys AGC	Азр ТТТ	Leu CAA	Gln TTG	Сув Сла	Leu GAT	Glu GCT	Asp 400 GÅG	Glu AAT	Leu	Gly ATC	Pro AGC	Leu AAT	Arg Atc	His AGA	Val GTA	Leu ACT	Asp GTT	Let GT/
аст са	A AGC n Ser	Lou AAA Lys	Lys AGC Ser	Asp TTT Phe	Leu CAA Gln	Gln TTG Leu	Cys GAA Glu	Leu GAT Asp	Glu GCT Ala	Asp 400 GÁG Glu	Glu AAT Asn	Leu TTC Phe	Gly ATC Ile	Pro AGC Ser	Leu AAT Asn	Arg ATC Ile	His AGA Arg	Val GTA Val	Leu ACT Thr	Asp GTT Val	Let GT/ Val
ACT CA Thr Gli	А АGC n Ser а Аас	Lou AAA Lys GT	Lys AGC Ser	Asp TTT Phe	Leu CAA Gln	Gln TTG Leu	Cys GAA Glu	Leu GAT Asp	Glu GCT Ala	Asp 400 GÁG Glu	Glu AAT Asn	Leu TTC Phe	Gly ATC Ile	Pro AGC Ser	Leu AAT Asn	Arg ATC Ile	His AGA Arg	Val GTA Val	Leu ACT Thr	Asp GTT Val	CT: Va
ACT CA Thr Gli Ana CT.	A AGC n Ser A AAG u Lys	Lou AAA Lys GT	Lys AGC Ser MAGG	Asp TTT Phe FGTT	Leu CAA Gln GCTT	Gln TTG Leu FATT	Cys GAA Glu TGCT/	Leu GAT Asp	Glu GCT Ala TGGA	Asp 400 GÅG Glu AATA	Glu AAT Asn	Leu TTC Phe	Gly ATC Ile AAGA	Pro AGC Ser	Leu AAT Asn	Arg ATC Ile	His AGA Arg	Val GTA Val	Leu ACT Thr	Asp GTT Val	CT: Va
ACT CA. Thr Gli AAA CT. Lys Les	A AGC n Ser A AAG u Lys	Lou AAA Lys GT	Lys AGC Ser MAGG	Asp TTT Phe FGTT	Leu CAA Gln GCTT	Gln TTG Leu FATT	Cys GAA Glu FOCT/	Leu GAT Asp	Glu GCT Ala TGGA	Asp 400 GÅG Glu AATA	Glu AAT Asn AAAT/	Leu TTC Phe AGAG	Gly ATC Ile AAGA	Pro AGC Ser	Leu AAT Asn	Arg ATC Ile	His AGA Arg	Val GTA Val	Leu ACT Thr	Asp GTT Val	CT: Va
ACT CA. Thr Gli AAA CT. Lys Les	A AGC n Ser A AAG u Lys	Lou AAA Lys GT	Lys AGC Ser MAGG	Asp TTT Phe FGTTC	Leu CAA Gln GCTT.	Gln TTG Leu FATT:	Cys GAA Glu FGCT/	Leu GAT Asp MATC	Glu GCT Ala TGGA AGTC: ~100	Asp 400 GAG Glu AATA ACTG 0 bp	Glu AAT ASD MAAT/ IGCA/	Leu TTC Phe AGAG	Gly ATC Ile AAGA	Pro AGC Ser	Leu AAT Asn	Arg ATC Ile	His AGA Arg AGTG	Val GTA Val GCTT	Leu ACT Thr SCCA	Asp GTT Val	GT: Va TGG
ACT CA Thr Gli AAA CT. Lys Le: CTTTGA	A AGC n Ser A AAG u Lys TGGGT	Lou AAA Lys GT	Lys AGC Ser AAGG	Asp TTT Phe FGTTC	Leu CAA Gln GCTT.	Gln TTG Leu FATT: ACCAJ GAAT	Cys GAA Glu FGCT/	Leu GAT Asp AATC TTAA CAGA	Glu GCT Ala TGGA AGTC -100 AGTG	Asp 400 GAG Glu AATA ACTG 0 bp TTCA	Glu AAT ASD GCAJ GCAJ	Leu TTC Phe AGAG	Gly ATC Ile AAGA	Pro AGC Ser ATG	Leu AAT Asn CATT CTCG	Arg ATC Ile TTTA	His AGA Arg AGTG	Val GTA Val GCTT CAAC	Leu ACT Thr GCCA	Asp GTT Val TTTC: GAGAG	GT/ Va FGG
ACT CA. Thr Gli AAA CT. Lys Les	A AGC n Ser A AAG u Lys TGGGT	Lou AAA Lys GT	Lys AGC Ser AAGG	Asp TTT Phe FGTTC	Leu CAA Gln GCTT.	Gln TTG Leu FATT: ACCAJ GAAT	Cys GAA Glu FGCT/	Leu GAT Asp AATC TTAA CAGA	Glu GCT Ala TGGA AGTC -100 AGTG	Asp 400 GAG Glu AATA ACTG 0 bp TTCA	Glu AAT ASD GCAJ	Leu TTC Phe AGAG	Gly ATC Ile AAGA	Pro AGC Ser ATG	Leu AAT Asn CATT CTCG	Arg ATC Ile TTTA	His AGA Arg AGTG	Val GTA Val GCTT CAAC	Leu ACT Thr GCCA	Asp GTT Val TTTC: GAGAG	GT/ Va FGG
ACT CA Thr Gli AAA CT. Lys Let CTTTGA TATTTT CAGTGT	A AGC n Ser A AAG u Lys TGGGT TGGGT TTATG TTATG TAAAA 450	Lou AAA Lys GT TCTG CTTT	Lys AGC Ser NAGG TGCA ACCA TGCC	Asp TTT Phe IGTTO ITTAC	Leu CAA Gln GCTT GTCA	GIN TTG Leu IATT GAAT CATAL	Cys GAA Glu FGCT/ AAGT FCTAG ATAAJ	Lou GAT Asp MATC TTAA CAGA MACA SAAT	Glu GCT Ala TGGA AGTC -100 AGTG TAAC	Asp 400 GÅG Glu AATA ACTG 0 bp TTCA CCAA ATAT	Glu AAT Asn IGCAJ IGCAJ IGCAJ	Leu TTC Phe AGAG AGTG FCCC	Gly ATC Ile AAGA AATCA ATCA ATCA	Pro AGC Sor AATGO AATGO PTTC/	Leu AAT Asn CATT CTCG MGAAJ	Arg ATC Ile TTTA TTGG TGG	HIS AGA Arg AGTG AGTG IAAT GAAT	Val GTA Val GCTTO CAAC GTTC.	Leu ACT Thr SCCA SCCA NATA	Asp GTT Val FTTC: GAGAC	
ACT CA. Thr Gli AAA CT. Lys Let CTTTGA CTTTGA TATTTT CAGTGT QGC TC	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TAAAA T GAC	Lou AAA Lys GT TCTG CTTT CTTT AAC	Lys AGC Ser AGCA TGCA ACCA ACCA	Asp TTT Pho FGTTC FTTAC FTTC FTTC	Leu CAA Gln GCTT GTCA GTCA GTCA GTCA GAG	GIN TTG Leu FATT: GAAT: GAAT: CATAI	Сув GAA Glu ГОСТ/ ПОСТ/ ПОСТ/ ГОСТ/ ПОСТ/	Lou GAT Asp MATC TTAA CAGA MACA SAAT/ TTC	Glu GCT Ala TGGA AGTC -100 AGTG TAAC AATT	Азр 400 GÅG Glu ААТА АСТС ⁶ 0 Бр ТТСА ССАА ⁶ АТАТ GAT	Glu AAT ASD IGCAJ IGCAJ IGCAJ IGCAAT GAG	Leu TTC Phe AGAG AGTG TCCC IGTG AGTA	Gly ATC 110 AAGA AATCA ATCA ATCA ACCA GCA	AGC Ser ATG	Leu AAT Asn CATT CTCG MGAAJ TAGCC	Arg ATC Ile TTTA TTTA TTGG TGT GTG	HIS AGA Arg AGTG AGTG IAAT GAAT GAC	Val GTA Val GCTT CAAC GTTC TTT	Leu ACT Thr SCCA SCCA NATA ATTC	Asp GTT Val FTTC: GAGAC FATT: AGG	
ACT CA Thr Gli AAA CT. Lys Let CTTTGA TATTTT CAGTGT	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TAAAA T GAC	Lou AAA Lys GT TCTG CTTT CTTT AAC	Lys AGC Ser AGCA TGCA ACCA ACCA	Asp TTT Pho FGTTC FTTAC FTTC FTTC	Leu CAA Gln GCTT GTCA GTCA GTCA GTCA GAG	GIN TTG Leu FATT: GAAT: GAAT: CATAI	Сув GAA Glu ГОСТ/ ПОСТ/ ПОСТ/ ГОСТ/ ПОСТ/	Lou GAT Asp MATC TTAA CAGA MACA SAAT/ TTC	Glu GCT Ala TGGA AGTC -100 AGTG TAAC AATT	Азр 400 GÅG Glu ААТА АСТС ⁶ 0 Бр ТТСА ССАА ⁶ АТАТ GAT	Glu AAT ASD IGCA IGCA IGCA IGCA IGCA GAG Glu	Leu TTC Phe AGAG AGTG TCCC IGTG AGTA	Gly ATC 110 AAGA AATCA ATCA ATCA ACCA GCA	AGC Ser ATG	Leu AAT Asn CATT CTCG MGAAA TAGCC	Arg ATC Ile TTTA TTTA TTGG TGT GTG	HIS AGA Arg AGTG AGTG IAAT GAAT GAC	Val GTA Val GCTT CAAC GTTC TTT	Leu ACT Thr SCCA SCCA NATA ATTC	Asp GTT Val FTTC: GAGAC FATT: AGG	
ACT CA. Thr Gli AAA CT. Lys Le CTTTGA CTTTGA TATTTT CAGTOT GCC TC Gly Be TGG AT.	A AGC n Ber A AAG u Lys TGGGT TTATG TTATG TTATG TAAAA 450 T GAC r Asp A GCC	Lou AAA Lys GT TCTG CTTT CTTT AAC AST	Lys AGC Sor AGCA FGCA ACCA ACCA TGCC TGC TGC	Asp TTT Pho FGTTO FTTAC FTTC FTTC FTTC FTTC FTTC FTTC FTT	Leu Gin Gin GCTT. GTCAJ GTCAJ GAG GAG GAG GAG GAG	GIN TTG Leu TATT GAAT CATAL CGT TGC Cys	Суя GAA Glu TOCT/ TOCT/ TCTAC ATGAA/ CAA Gln ATC	Lou GAT Asp MATC FTAN CAGA MACA SAAT TTC Pho TCA	Glu GCT Ala TGGA. ~100 AGTC TAACC GAT ASP ACA	Asp 400 GÅG Glu AATA ACTG 0 bp TTCA CCAA CCAA ATAT Asp AGC	Glu AAT ASD IGCAJ IGCAJ IGCAJ IGCAJ GAG GLU S500 CCT	Leu TTC Phe AGAG AGTG TCCC. TCCC. TCTG Ser CAA	Gly ATC Ile AAGA AATCA ATCA ATCA AGCT Ale	AGC Sor ATGO ATGO TTC: ACT Thr	Leu AAT Asn CATT CTCG MGAAJ TAGCO GTG Val	Arg ATC Ile TTTA TTGG TTGG TTAC	His AGA Arg AGTG TAAT GAAT GAAT GAC Asp	Val GTA Val GCTT CAAC GTTC TTT Phe	Leu ACT Thr SCCA SCCA RCTG AATA ATTC Leu	Asp GTT Val FTTC: GAGAG FATT: AGG Arg	Let GT/ Val FGGT GGT FGGT
ACT CA Thr Gli AAA CT. Lys Lec CTTTGA CTTTGA TATTTT CAGTGT GGC TC Gly Be	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TTATG T GAC r Asp A GCC Ala	Leu AAA Lys GTJ TCTG CTTT TAAAC Asn TTC Phe	Lys AGC Sor AGCA FGCA ACCA ACCA TGCC TGC TGC	Asp TTT Pho FGTTO FTTAC FTTC FTTC FTTC FTTC FTTC FTTC FTT	Leu Gin Gin GCTT. GTCAJ GTCAJ GAG GAG GAG GAG GAG	GIN TTG Leu TATT GAAT CATAL CGT TGC Cys	Суя GAA Glu TOCT/ TOCT/ TCTAC ATGAA/ CAA Gln ATC	Lou GAT Asp MATC FTAN CAGA MACA SAAT TTC Pho TCA	Glu GCT Ala TGGA. ~100 AGTC TAACC GAT ASP ACA	Asp 400 GÅG Glu AATA ACTG 0 bp TTCA CCAA CCAA ATAT Asp AGC	Glu AAT ASD IGCAJ IGCAJ IGCAJ IGCAJ GAG GLU S500 CCT	Leu TTC Phe AGAG AGTG TCCC. TCCC. TCTG Ser CAA	Gly ATC Ile AAGA AATCA ATCA ATCA AGCT Ale	Pro AGC Ser AATG C AATG ITTC: ACT Thr	Leu AAT Asn CATT CTCG MGAAJ TAGCO GTG Val	Arg ATC Ile TTTA TTGG TTGG TTAC	His AGA Arg AGTG TAAT GAAT GAAT GAC Asp	Val GTA Val GCTT CAAC GTTC TTT Phe	Leu ACT Thr SCCA SCCA RCTG AATA ATTC Leu	Asp GTT Val FTTC: GAGAG FATT: AGG Arg	Let GT/ Val FGGT GGT FGGT
ACT CA. Thr Gli AAA CT. Lys Le CTTTGA CTTTGA TATTTT CAGTOT GCC TC Gly Be TGG AT.	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TAAAA 450 T AASP A GCC A AGC	Lou AAA Lys GT. TCTG CTTT, TOAA Asn TTC Phe 90	Lys AGC Ser AAGG TGCA ACA Thr TGT Cys	Asp TTT Phe TGTTC ITTAC ITTAC ITTG ITTAC CAA Gln ITTAAC	Leu CAA Gin GCTT GCTT ATATA AGCT GAG Glu AGC Ser	Gln TTG Leu TATT: ACCAJ GAAT: CATAJ CCGT: TGC Cys ATC 11e	Cys GAA Glu FOCT/ AAGT: FCTAG ATAAJ CAA Gln ATC Ile	Leu GAT Asp AATC TTAA CAGAA AACA TTC Phe TCA Ser	Glu GCT Ala TGGA AGTC. ~100 AGTC AGTC AACT ASP ACA Thr	Asp 400 GAG Glu AATA ACTG 0 bp TTCA CCAA ATATT GAT Asp ACC Ser	Glu AAT ASN GCAA CTGCA TAAT TAAT TAAT CAA GAG Glu S50 CCT Pro	Leu TTC Phe AGAG AGTG TCCC TCCC TCTG Ser CAA Gln	Gly ATC Ile AAGA AATCA ATCA ATCA ATCA ACCT. GCA Ale TA	Pro AGC Ser AATGO C AATGO TTTC/ ACCTT Thr 650	Leu AAT Asn CATT CTCGT AGAAJ FAGCC GTG Val	Arg ATC Ile TTTA TTTA TTGG TTGG TTAC CTAC	His AGA Arg AGTG TAAT GAAT AATT GAAT ASP CTGC	Val GTA Val GCTTU CAAC GTTC. TTTTTTTTTT Phe TTAC.	Leu ACT Thr SCCA TCTG AATA CTG Leu AACM	Asp GTT Val ITTC: GAGAGAG FATT: AGG Arg CATAJ	Let GTJ Val FGGT CCTI FAAC AGA
ACT CA Thr GLI AAA CT. Lys Let CTTTGA CATTTGA TATTTT CAGTOT GGC TC GIY Se TGG AT. TTP II: TCTCTA GATCTT	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TAAAG A GCC A AGC A	Lou AAA Lys GTJ TCTG CTTT. TAAA ASn TTC Phe 00 TTAA	Lys AGC Ser TGCA TGCA ACCA TGCC ACA Thr TGT Cys ATAT	Asp TTT Phe IGTTA ITTAA ITTG IAAA Gln TTAA Gln ITAAA Gln ITAAA	Leu CAA GIN GCIT GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ	GIN TTG Leu FATT ACCAJ GAAT CATAJ CCATA TGC CYs ATC LIC IIC	Cys GAA Glu TOCT/ TOCT/ TCTAC TCTAC TCTAC TCTAC CAA Gln ATC Ile TATT CGGC	Leu GAT Asp TTAAJ CAGA AACA DAACA TTC Phe TCA Ser TTTG TTTG	Glu GCT Ala TGGA AGTC. ~100 AGTG TAACO AATT. ASP ACA Thr GATG ACA	Asp 400 GÅG GÅU AATA ACTG 0 bp TTCA CCCAA ATAT GAT ASP AGC Ser TATT	Glu AAT ASD GCA GCA GLU STGT PTO STTT/	Leu TTC Phe AGAG AGTG TCCC. TGTG. AGTA Ser CAA Ser CAA CIA	Gly ATC Ile AAGA AATCA ATCA ATCA ATCA ATCA ATCA ATC	Pro AGC Ser MATGO C MATGO C MATGO C ACT Thr ACTA Thr 650 ProT/ CTGA	Leu AAT Asn CATT CTCG CTCG Val TGTAC AAACT/	Arg ATC Ile TTA TTA TTA TTA TTA CTA CTA CTA TTA CTA TTA CTA	His AGA Arg AGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG	Val GTA Val GCTTV CAAC GTTC. TTTAT. Phe TTAC. FCAG	Leu ACT Thr SCCA SCCA NCTG CTG Leu AACA ATGA CTGA	Asp GTT Val EFFC SAGAGA FATT AGG Arg CATAJ	
ACT CA Thr GLI AAA CT. Lys Let CTTTGA CATTTGA TATTTT CAGTOT GGC TC GIY Se TGG AT. TTP II: TCTCTA GATCTT	A AGC n Ser A AAG u Lys TGGGT TTATG TTATG TAAAG A GCC A AGC A	Lou AAA Lys GTJ TCTG CTTT. TAAA ASn TTC Phe 00 TTAA	Lys AGC Ser TGCA TGCA ACCA TGCC ACA Thr TGT Cys ATAT	Asp TTT Phe IGTTA ITTAA ITTG IAAA Gln TTAA Gln ITAAA Gln ITAAA	Leu CAA GIN GCIT GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ GTCAJ	GIN TTG Leu TATT ACCAJ GAAT CATAJ CCATA TGC CYs ATC LIC LIC LCAAC	Cys GAA Glu TOCT/ TOCT/ TCTAC TCTAC TCTAC TCTAC CAA Gln ATC Ile TATT CGGC	Leu GAT Asp TTAAJ CAGA AACA DAACA TTC Phe TCA Ser TTTG TTTG	Glu GCT Ala TGGA AGTC. ~100 AGTG TAACO AATT. ASP ACA Thr GATG ACA	Asp 400 GÅG GÅU AATA ACTG 0 bp TTCA CCCAA ATAT GAT ASP AGC Ser TATT	Glu AAT ASD GCA GCA GLU STGT PTO STTT/	Leu TTC Phe AGAG AGTG TCCC. TGTG. AGTA Ser CAA Ser CAA CIA	Gly ATC Ile AAGA AATCA ATCA ATCA ATCA ATCA ATCA ATC	Pro AGC Ser MATGO C MATGO C MATGO C ACT Thr ACTA Thr 650 ProT/ CTGA	Leu AAT Asn CATT CTCG CTCG Val TGTAC AAACT/	Arg ATC Ile TTA TTA TTA TTA TTA CTA CTA CTA TTA CTA TTA CTA	His AGA Arg AGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG GAGTG	Val GTA Val GCTTV CAAC GTTC. TTTAT. Phe TTAC. FCAG	Leu ACT Thr SCCA SCCA NCTG CTG Leu AACA ATGA CTGA	Asp GTT Val EFFC SAGAGA FATT AGG Arg CATAJ	
ACT CA Thr Gli AAA CT. Lys Le CTTTGA' CAGTGT CAGTGT CAGTGT CAGTGT CAGTGT TGG AT. TTD II: CTCTA	A AGCC n Ser TGGGT TGGGT TTATG TAAAA TGACA A SC A SC A AGCC A AGCC A AGCC A AGCC A AGCC A AGCC A AGC A CONTACT A AGCC A AGCC A AGCC A AGC A AGCC A AGC A AGCC A AGCC	Lou AAA Lys GT. TCTG CTTT. TAAA AAC AAS TTC Phe 90 TTAAA ATTC GTAG	Lys AGC Ser AAGG FGCA ACCA TGC TGC TGC TGC Cys ACA	Asp TTT Pho TGTTV TTTA TTTG TTTA CAA Gln TTT Pho CAA Gln TTA CAA	Leu CAA Gln GCTT TCAJ TCAJ TCAJ GAC GAC GAC Ser CTTTJ AGCCC	GIN TTG Leu TATT GAAT CATA CATA ACCAU GAAT CCT Ile ATC CY3 ATC CY3 Ile AATC S00 AATC	Cys GAA Glu FGCT/ AAGT FCTAA ATAAJ CAA Gln ATC Ile FATT SGCC	Leu GAT Asp TTAAJ TTAAJ CAGAA CAGAA CAGAA TTAA TTCC Phe TCAA Ser TTTG TCAAJ	Glu GCT Ala TGGA AGTC. ~100 AGTC TAACC AATT. GATT GATG AAATC GATTO GATTO	Авр 400 GÅG Glu ААТАА АСТСС 0 bp ТТСАА СССАА АТАТ СССАА АСТСС БОГ ТТСАА СССАА АСТСС СССАА АСТСС СССАА ССССАА СССССССАА ССССАА ССССАА ССССАА ССССАА ССССАА СССССАА СССССССС	Glu AAT ASD IGCAJ IGCAJ IGCAJ IGCAJ IGCAJ IGCAA IGCAAJ IGCAAJ IGCAAJ IGCAAJ IGCAAJ IGCAAJ IGCAAJ IGCAAJ IGCA IGCAJ I IGO	Lou TTC Phe AGAG AGTG TCCC. TCCC. TCGG Ser CAA Ser CAA Strata	Gly ATC Ile AAGA AATCA ATCA ATCA ATCA ACCT TA TA AGCT	Pro AGC Ser AATG C AATG TTC: ACT Thr ACTA CAGA CAGA	Leu AAT Asn CATT CATT CATT CATCG CTCG CTCG CTCG Val IGTAC AACT/ AATT/	Arg ATC Ile TTA TTG TTAC GTG Val CTAC Val	His AGA Arg Agro Garg Garg Asp TGC TGC	Val GTA Val GCTTV CAAC GTTC. TTTT Phe TTAC. B TTAC. B TTAC.	Leu ACT Thr SCCA SCCA CTG Leu AATA ATTC Leu AACA ATGA SAAT	Asp GTT Val TTTC' GAGA ASG CATAJ CATAJ CATAJ	

GATOSCTTGTGGGAAAAGATCTCCTCTCCAGGGAGCTAACATCAGCTCAGAGTTTACTCAAGAATTC

Fig. 3. Nucleotide sequence of mouse IL-2 gene. Four exons are framed. Numbers refer to nucleotide positions of exons from the presumed cap site. Dots and open circles indicate TATA box and poly A additional signals, respectively.



Fig. 4. Comparison of 5'-flanking regions of mouse and human IL-2 genes. In aligning the sequences for both genes, gaps were introduced to maximize homology. Dots indicate identical nucleotide sequences. Number 1 indicates putative transcription initiation site. Inverted repeats are indicated by bars and dashed lines. TATA box is framed.

of both genes is illustrated in Fig. 4. The nucleotide sequence homology from the TATA box to -470 is 85%. The highest region of homology was observed between the TATA box and position -97 (60 bp matches out of 64 bp) of both genes.

DISCUSSION

We have isolated recombinant clones for mouse IL-2 gene from a phage Charon 4A/mouse genomic DNA library and determined the entire sequence of the gene except for the sequence of the internal portion of the second and third introns. The mouse IL-2 cDNA sequence was aligned with the genomic sequence and both sequences matched completely each other. The unusual CAG repeats encoding 12 glutamines which was found previously in the cloned cDNA was shown to be present also in the chromosomal gene. This finding further excludes the possibility that the unique repeat is generated by artifacts during the cDNA cloning process. Although we can not rule out the possibility for the deletion of this sequence in the human IL-2 gene, it is more likely that the CAG repeat has been generated in the mouse genome rather The repeat could have been generated either by a recently. direct insertion of the sequence or by the duplication after insertion of a unit sequence.

Organization of the mouse IL-2 gene resembles to that of the human gene (Fig. 2., ref. 9). There seems to be little sequence homology between corresponding introns of mouse and human IL-2 genes, except for the intron-exon junctions part of which is thought to be necessary for the RNA splicing (17, 18). Dissimilarity of the intron sequences among the genes which are derived from a common ancestor has been reported in other genes (19, 20). In spite of the divergence in sequence of introns, the size and position of the introns are very similar between the murine and human IL-2 genes.

Of particular interests is the presence of highly conserved sequences in the 5' -flanking region of the human and mouse IL-2 gene (Fig. 4). Whereas the coding region shows nucleotide sequence homology of 72% between the two genes, the 5' upstream region spanning about 500 bp (Fig. 4) shows 85% homology which was readily detectable by the blotting analysis (Fig. 1, lane 1-4). Since we have not yet determined the nucleotide sequence further upstream of the mouse gene, we do not know whether or not this similarity extends further. It is likely that such sequences are involved in the controlled expression of the IL-2 genes in activated T-lymphocytes. Work is in progress to identify such DNA sequences by introducing the cloned genes into various lymphocytic cell lines. Our preliminary results indicate that the 5 '-flanking sequence of the human IL-2 gene mediates mitogen induced expression of the gene in T-lymphocytic cells (Fujita & Taniguchi, unpublished observation).

ACKNOWLEDGEMENTS

We thank Dr. T. Honjo for mouse gene library. We are also indepted to Ms. M. Nagatsuka for typing the manuscript. This work was supported in part by Grant-in-Aid for Special Project Research, Cancer-Bioscience from the Ministry of Education, Science and Culture, Japan.

§To whom correspondence should be addressed

*Present address: Department of Microbiology, School of Medicine, Chiba University, Chiba 280, Japan

+ Present address: Central Research Laboratory, Ajinomoto Co.Inc., Totsuka-ku, Yokohama 244, Japan

REFERENCES

- Morgan, D. A., Ruscetti, F. W. and Gallo, R. (1976) Science, 193, 1007-1008.
- Gillis, S. Ferm, M. M., Ou, W. and Smith, K. (1978) J. Immunol., 120, 2023-2027.
- Chen, B. M. and Di Sabato, G. (1976) Cell. Immunol., 22, 211-224.
- Henney, C. S., Kuribayashi, K., Kern, D. E. and Gillis, S. (1981) Nature, 291, 335-338.
- 5. Wagner, H., Hardt, C., Heeg, K., Rollinghoff, M. and Pfizenmaier, K. (1980) Nature, 284, 278-280.
- 6. Farrar, J. J., Benjamin, W. R., Hilfikr M. L., Howard, M., Farrar,W. L. and Fuller-Farrar, J. (1982) Immunol. Rev. 63, 129-166.
- Pearlstein, K., Palladıno, M. A., Welte, K. and Vilcek, J. (1983) Cell. Immunol., 80, 1-9.
- Taniguchi, T., Matsui, H., Fujita, T., Takaoka, C., Kashima, N., Yoshimoto, R. and Hamuro, J. (1983) Nature 302, 305-310.
- 9. Fujita, T., Takaoka, C., Matsui, H. and Taniguchi T. (1983) Proc. Natl. Acad. Sci. USA 80, 7437-7411.
- 10. Ohno, S. and Taniguchi, T. (1982) Nucleic Acids Res. 10, 967-977.
- 11. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- 12. Benton, W. D. and Davis, R. W. (1977) Science 196, 180-182.
- 13. Ohno, S. and Taniguchi, T. (1981) Proc. Natl. Acad. Sci. USA 78, 5305-5309.
- 14. Maxam, A. M. and Gilbert, W. (1980) Meth. Enzymol. 65, 560-580.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- 16. Gillis, S., Scheid, M. and Watson, J. D.(1980) J. Immunol. ,125, 2570-2580.
- 17. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P.(1978) Proc. Natl. Acad. Sci. USA, 75, 4853-4857.
- 18. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980) Nature, 283, 220-224.
- 19. Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissmann, C.(1979) Science, 206, 337-344.
- Searle, P. F., Davison, B. L., Stuart, G. W., Wilkie, T. M., Norstedt, G. and Palmiter, R. D. (1984) Mol. Cell. Biol. 4, 1221-1230.