

# Gene structures of the $\alpha$ subunits of human IL-3 and granulocyte-macrophage colony-stimulating factor receptors: Comparison with the cytokine receptor superfamily

Hiroshi Kosugi, MD,<sup>a, b</sup> Yasushi Nakagawa, MD,<sup>a</sup> Ken-ichi Arai, MD, PhD,<sup>a</sup> and Takashi Yokota, PhD<sup>a</sup> Tokyo and Nagoya, Japan

*Recently, many genes encoding the members of the cytokine receptor superfamily (CRSF), which have common structural features, have been characterized. Analyses on the structures of the genes encoding the  $\alpha$  subunits of human IL-3 (hIL-3R $\alpha$ ) and granulocyte-macrophage colony-stimulating factor receptors (hGMR $\alpha$ ) revealed that they have the structural features common to all members of the CRSF (i.e., conservation of the intron phase pattern as "1-2-1-0-1" rule in the fibronectin type III domains located in extracellular segments of type I cytokine receptor subunits. This finding led us to propose a possible model for gene evolution for the CRSF. We pointed out that the CRSF genes derived from a putative common ancestral gene. In addition to these common features, we found an additional intron that is unique to the IL-3R $\alpha$  and the GMR $\alpha$  genes. This additional intron suggests that the IL-3R $\alpha$  and the GMR $\alpha$  genes evolved closely in the evolution process of the CRSF genes. This evidence and results of recent studies on the evolution of mammalian X chromosome make it tempting to speculate that a putative common ancestral gene of the subfamily including IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  emerged in an autosome at least before the divergence of marsupials and eutherian mammals, early in the 200 million-year history of mammals. (J ALLERGY CLIN IMMUNOL 1995;96:1115-25.)*

**Key words:** Cytokine receptor, gene structure, gene evolution, IL-3, GM-CSF, cytokine receptor superfamily

The receptors for IL-3, granulocyte-macrophage colony-stimulating factor (GM-CSF), and IL-5 consist of two subunits, the  $\alpha$  subunit specific to each and the common  $\beta$  subunit.<sup>1</sup> Pleiotropy and redundancy are the features of the cytokine action, which are to some extent accounted for by receptor subunit sharing and signal transduction molecule sharing in the downstream, such as Ras, MAP kinase, and JAK kinases.<sup>2</sup>

From the protein structure, cytokine receptors are classified into four classes of the cytokine receptor superfamily (CRSF).<sup>3</sup> The  $\alpha$  subunits

## Abbreviations used

|         |                                                  |
|---------|--------------------------------------------------|
| CRSF:   | Cytokine receptor superfamily                    |
| Fn3:    | Fibronectin type III                             |
| GM-CSF: | Granulocyte-macrophage colony-stimulating factor |
| GMR:    | GM-CSF receptor                                  |

of the receptors for IL-3 (IL-3R $\alpha$ ), GM-CSF (GMR $\alpha$ ), and IL-5 (IL-5R $\alpha$ ) are classified together with other hemopoietic cytokine receptors as type I because of the common structural feature in the extracellular segment, that is, each tandem array of approximately 100 amino acids of domains similar to fibronectin type III (Fn3) domain, "N domain," which is located in its N terminus including four conserved cysteine residues and "C domain," located in its C terminus including the "WSxWS" motif.

Recently, many genes encoding the CRSF members have been revealed. Although resemblance in

From <sup>a</sup>Department of Molecular and Developmental Biology, The Institute of Medical Science, The University of Tokyo; and <sup>b</sup>The First Department of Internal Medicine, Nagoya University School of Medicine, Nagoya.

Reprint requests: Takashi Yokota, PhD, Department of Molecular and Developmental Biology, The Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan.

Copyright © 1995 by Mosby-Year Book, Inc.  
0091-6749/95 \$5.00 + 0 1/0/68342

**TABLE I.** Genomic libraries used for screening for hIL-3R $\alpha$  and hGMR $\alpha$  genes

|           | Chromosome | Vector          | Insert size | Digestion               |
|-----------|------------|-----------------|-------------|-------------------------|
| Library A | Total      | EMBL 3          | 10-14 kb    | <i>Bam</i> HI, complete |
| Library B | Total      | $\lambda$ gtWES | 2-13 kb     | <i>Eco</i> RI, complete |
| Library C | X          | Charon 35       | 10-21 kb    | <i>Sau</i> 3AI, partial |
| Library D | Y          | Charon 40       | 9-24 kb     | <i>Mbo</i> I, partial   |
| Library E | Total      | pAd10SacBII     | 85-100kb    | <i>Sau</i> 3AI, partial |

genomic organization of the CRSF members has been indicated, no comprehensive picture of these common features of the gene structures of CRSF members has been presented. Through the analyses of the structures of the human IL-3R $\alpha$  (hIL-3R $\alpha$ ) and the human GMR $\alpha$  (hGMR $\alpha$ ) genes, we pointed out the conservation of the gene structures of the CRSF members, which led us to propose a model for the evolution of the CRSF.<sup>4,5</sup> On the basis of these analyses, we discuss a possible pathway for gene evolution of the CRSF members.

## METHODS

### Genomic cloning for hIL-3R $\alpha$ and hGMR $\alpha$

We screened five genomic libraries, in total, in order to isolate genomic clones for hIL-3R $\alpha$  and hGMR $\alpha$  (Table I). Library A was prepared by ligating EMBL3 lambda phage vector (Stratagene, La Jolla, Calif.) and human genomic DNA extracted from placenta of healthy volunteers and, thereafter, packaged them with GIGA PACK II GOLD (Stratagene). Library B was obtained from the Japanese Cancer Research Resources Bank (JCRB, Tokyo, Japan L1 105). Libraries C and D were from the American Type Culture Collection (ATCC, Rockville, Md., 57750 and 57780, respectively). Several fragments excised from the full-length complementary DNAs of the hIL-3R $\alpha$ <sup>6</sup> and hGMR $\alpha$ <sup>7</sup> were used as probes for plaque hybridization. For the hIL-3R $\alpha$  gene, we screened a total of  $1.0 \times 10^5$ ,  $1.3 \times 10^6$ , and  $1.5 \times 10^6$  clones of libraries A, C, and D, respectively. We isolated clones 3A1 and 25 from library A; clones 3A4, 5, 7, 8, 10, and 11 from library B; and clone 3A17 from library C. Because the contig of the lambda phage clones was found to lack the region spanning the exon 9, we obtained P1 phage clones 2057, 2058, and 2059, using polymerase chain reaction screening from Genome Systems Inc. (USA). For the hGMR $\alpha$  gene, we screened a total of  $1.8 \times 10^5$ ,  $3.2 \times 10^5$ , and  $3.2 \times 10^5$  clones of libraries B, C, and D, respectively. We isolated clones GA1 from library B, nine clones (GA3 to GA11) from library C, and clones GY5 and GZ12 of 260 positive clones from library D.

After subcloning these genomic clones into pBlue-script KS (+), (-), SK (-), or pUC19, restriction endonuclease cleavage maps (shown in Fig. 1, A and B) were determined by digestion with *Bam*HI, *Eco*RI, *Hind*III, *Xho*I, *Sac*I, *Eco*RV, and *Sal*I and polymerase

chain reaction with universal primers or exon-specific primers. These genomic clones cover the entire coding sequences of hIL-3R $\alpha$  and hGMR $\alpha$ . These restriction maps were consistent with the results from genomic Southern blot analyses (Fig. 2).

### Exon-intron organizations of hIL-3R $\alpha$ and hGMR $\alpha$ genes

Exon-intron boundaries (Fig. 3, A and B) for each of the genes were sequenced in a 373A DNA sequencer (ABI, Applied Biosystems Inc., Foster City, Calif.). The hIL-3R $\alpha$  and hGMR $\alpha$  genes span approximately 40 kb and 44 kb, respectively, and consist of 12 and 13 exons, respectively. All the exon-intron boundaries possess the splicing junction consensus sequences (GT-AG rule).<sup>8</sup> Although we determined the transcription start site of the messenger RNA for hGMR $\alpha$  by primer extension analysis and 5' RACE, we could not detect the distinct cDNA and of hIL-3R $\alpha$  using the two procedures described above. This may be due to the high GC content of the putative 5' untranslated region of the hIL-3R $\alpha$  gene. For example, we have sequenced 580 bp of the 5' flanking region from the reported cDNA end and found the SINES homologous region<sup>9</sup> with a GC content of more than 80%. This sequence would interfere with the reverse transcription used in the two procedures described above. Although we cannot exclude the possibility that the hIL-3R $\alpha$  gene contains one more exon in its 5' end, structures of hIL-3R $\alpha$  and hGMR $\alpha$  genes are the most similar among the CRSF members for which gene structures are known,<sup>10-17</sup> at least in the coding region (Fig. 4).

## RESULTS

### Sequential intron phase arrangement in the extracellular segment is conserved according to the "1-2-1-0-1 rule" among CRSF members

We compared the genomic structures of CRSF members, including those of hIL-3R $\alpha$  and hGMR $\alpha$  genes (Fig. 4, A), focusing especially on intron phase in the coding region. Intron phases are defined as follows.<sup>18</sup> Phase 0 introns interrupt the reading frame between codons. Phase 1 and 2 introns interrupt codons after the first and second nucleotides of codons, respectively.

We found the common feature in intron phases

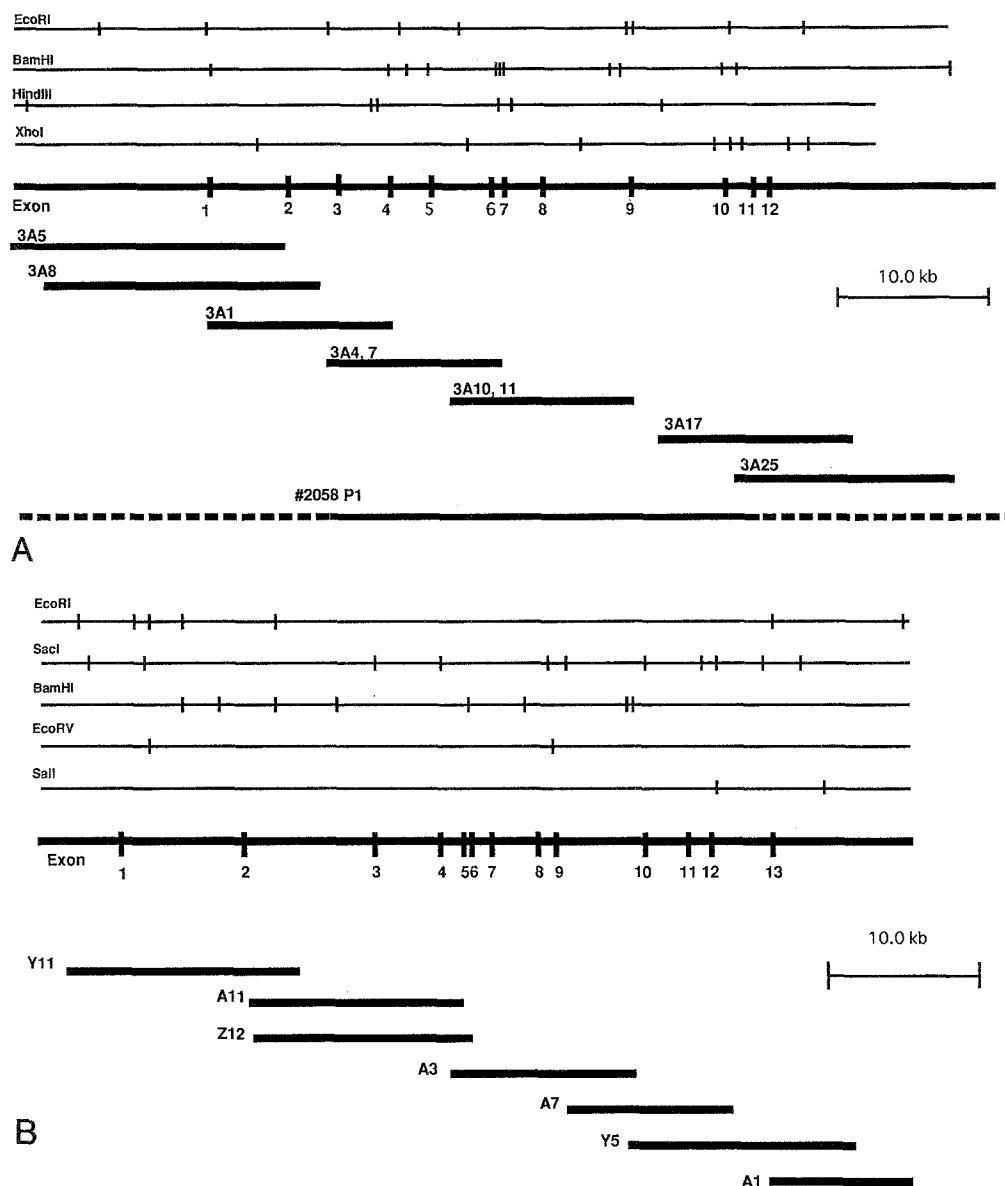


**FIG. 1.** Fine restriction endonuclease cleavage maps of hIL-3R $\alpha$  and hGMR $\alpha$  genes. **A**, Map of the hIL-3R $\alpha$  gene (5), which spans approximately 40 kb, including 12 exons. **B**, Map of the hGMR $\alpha$  gene (4), which spans approximately 44 kb, including 13 exons. Two types of isoforms of hGMR $\alpha$  reported (4). Soluble form of hGMR $\alpha$ <sup>52, 53</sup> is possible when exon 11 is spliced out to result in frameshifting and premature termination of translation. GMR.B<sup>54</sup> results in, containing different cytoplasmic region from the original form when exon 13' is not skipped.

in the extracellular segments of CRSF members. From the protein structure, the CRSF members have Fn3-like domains in their extracellular segments.<sup>19</sup> All of the boundaries of Fn3-like domains of the CRSF members are defined by phase 1 introns without exception. These Fn3-like domains are roughly halved intervening introns, of phase 2 in the N domain and of phase 0 in the C domain. Together with these two common features they

result in the sequential intron phases arranged as 1-2-1-0-1 (Fig. 4, B).

This conservation of the sequential arrangement of intron phases within Fn3-like domains should be related to gene evolution of CRSF members. However, many proteins, in organisms ranging from bacteria to vertebrates, other than cytokine receptors are known to contain Fn3-like structures.<sup>20</sup> All of their Fn3-like domains are surrounded by in-

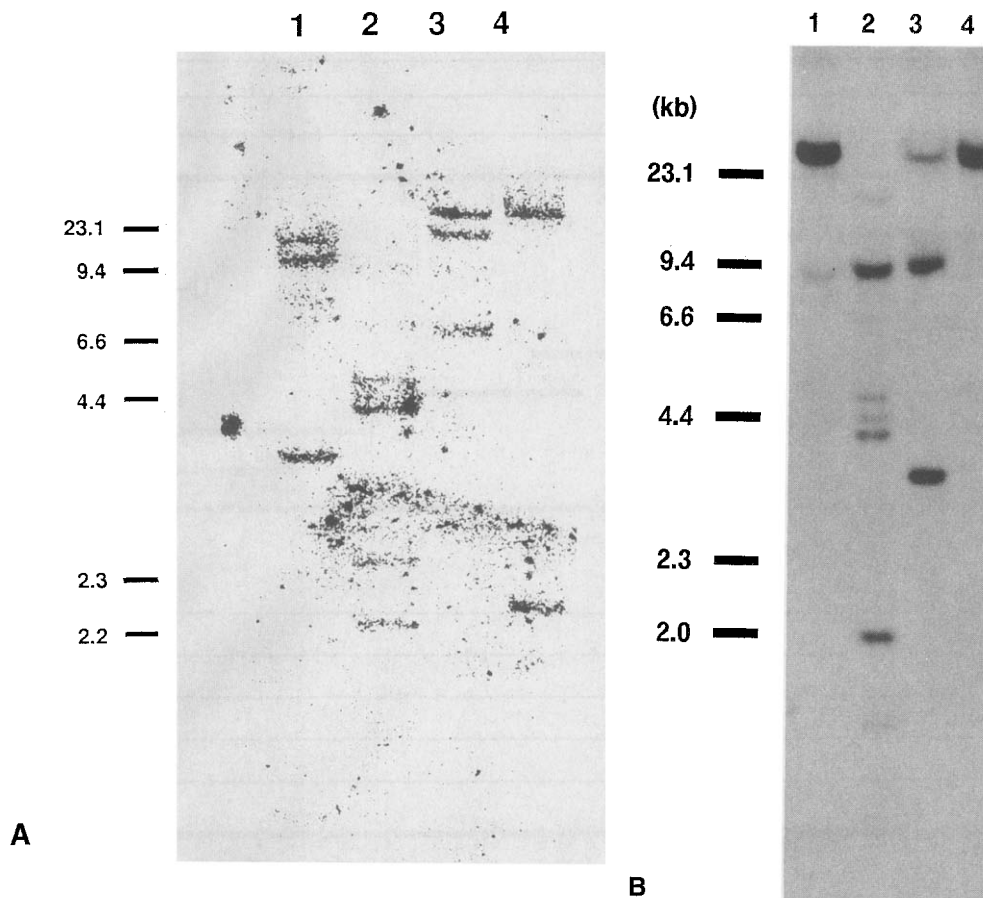


**FIG. 2.** Genomic Southern blot analysis for the hIL-3R $\alpha$  (A) and the hGMR $\alpha$  (B) genes. DNA size marker is shown in kilobases. A, Ten micrograms of human placental DNA was digested with *Bam*HI (lane 1), *Eco*RI (lane 2), *Hind*III (lane 3), and *Xho*I (lane 4) and electrophoresed on a 0.8% agarose gel. The DNA was then transferred to a nylon filter (Hybond N; Amersham International, Amersham, U.K.) and was hybridized with the full-length cDNA of the hIL-3R $\alpha$  gene. B, *Eco*RI (lane 1), *Sac*I (lane 2), *Bam*HI (lane 3), and *Eco*RV (lane 4). The region +188 to +1432 of hGMR $\alpha$  cDNA was used as a probe.

trons of phase 1, and most of them are separated by intervening introns of phase 0 or 1. In this way the Fn3-like domain, as a protein module, is supposed to be used in the exon-shuffling mechanism for gene evolution of many proteins. It should be noted that exons surrounded by introns of the same phase can be inserted or duplicated without frameshifting of the neighboring coding region. In addition to these studies of the Fn3-like domain, we found that the intervening intron of phase 2, observed in the N domain, is a unique usage, limited in the case of the CRSF to a variety of Fn3-like domains known in vertebrates. Indeed, usage of the phase 2 intron is rare according to recent exon-shuffling theory.<sup>21</sup> Based on these backgrounds, conservation of sequential arrangement of intron phases of Fn3-like domains of the CRSF can be regarded as the 1-2-1-0-1 rule.

#### Common intron phase is also seen in the cytoplasmic segment of the CRSF

We also found that gene structures of the cytoplasmic segment are similar to each other and that a common intron phase is observed in the cytoplasmic segment of CRSF members. As shown in Fig. 4, A, most of the cytoplasmic segment of type I CRSF consists of two exons, the exon neighboring the transmembrane region and the exon corresponding to the last, and the majority, of the cytoplasmic region. In the cases of the mIL-4R and AIC2A/B genes, cytoplasmic segments consist of three exons, which seems to be the result of insertion of an additional intron of phase 0 into the mIL-4R and AIC2A/B genes. In all of the genes of type I CRSF members available, the exon proximal to the transmembrane region is always surrounded by phase 2 (5' side) and phase 0 (3' side) introns.

**A**

| Exon No. | size(bp) | codon phase | Intron donor | size(kb) | Acceptor           |
|----------|----------|-------------|--------------|----------|--------------------|
| 1        | 108      |             | TTAAG        | 5.3      | tctcctccag CAGGC   |
| 2        | 102      | 1           | G GAA G      | 3.0      | accgtttttag AT CCA |
| 3        | 129      | 0           | TG CCG       | 2.9      | tctctcttag GCA GT  |
| 4        | 145      | 1           | G AAC A      | 2.5      | gtgaaccag GT GGG   |
| 5        | 133      | 2           | GCC AA       | 4.9      | cttaccgcag C AGG C |
| 6        | 85       | 1           | G ATT G      | 1.8      | tcgttgcag AG ATA   |
| 7        | 126      | 0           | AA AAG       | 2.0      | tgtgttcag AGA AT   |
| 8        | 27       | 0           | AA CAG       | 6.0      | caaaccacag GTC AG  |
| 9        | 115      | 1           | C TTC G      | 6.4      | ttaccctag AG TGC   |
| 10       | 106      | 2           | AGA AG       | 2.0      | ttcctccag G TAT C  |
| 11       | 82       | 0           | AG CTG       | 1.1      | gtctctgcag GTG GT  |
| 12       | 252      |             |              |          |                    |

**B**

| Exon No. | size(bp) | codon phase | Intron donor | size(kb) | Acceptor         |
|----------|----------|-------------|--------------|----------|------------------|
| 1        | 81       |             | TGGGG        | 7.3      | tctcacag TTTAC   |
| 2        | 88       |             | TGTCG        | 7.4      | tcctgcag CTCTT   |
| 3        | 102      | 1           | A TCG G      | 3.7      | ccttgcag AT CTG  |
| 4        | 143      | 0           | CC AGG       | 2.7      | ttcttcag CTC AG  |
| 5        | 124      | 1           | T TCA G      | 0.1      | cttccag GA AGG   |
| 6        | 130      | 2           | TCA AA       | 1.5      | actatcag G AGA A |
| 7        | 173      | 1           | A ATA G      | 3.8      | cggtacag AA CGA  |
| 8        | 134      | 0           | GA AAG       | 1.0      | gtttctag AAT AC  |
| 9        | 30       | 0           | TA CTG       | 5.1      | ctccaaag ATT AA  |
| 10       | 136      | 1           | A TTT G      | 3.0      | tactccag GT TCT  |
| 11       | 97       | 2           | AAA AG       | 1.4      | tgtctcag G TTC C |
| 12       | 82       | 0           | AC GAG       | 4.2      | tttctcag ATC AT  |
| 13       | 533      |             |              |          |                  |

FIG. 3. Exon-intron boundaries of the hIL-3R $\alpha$  (A) and the hGMR $\alpha$  (B) genes.

This exon, in the case of gp130, corresponds to the region including the "box 1," which is believed to be conserved among CRSF members and essential for IL-6-induced cell proliferation.<sup>22</sup>

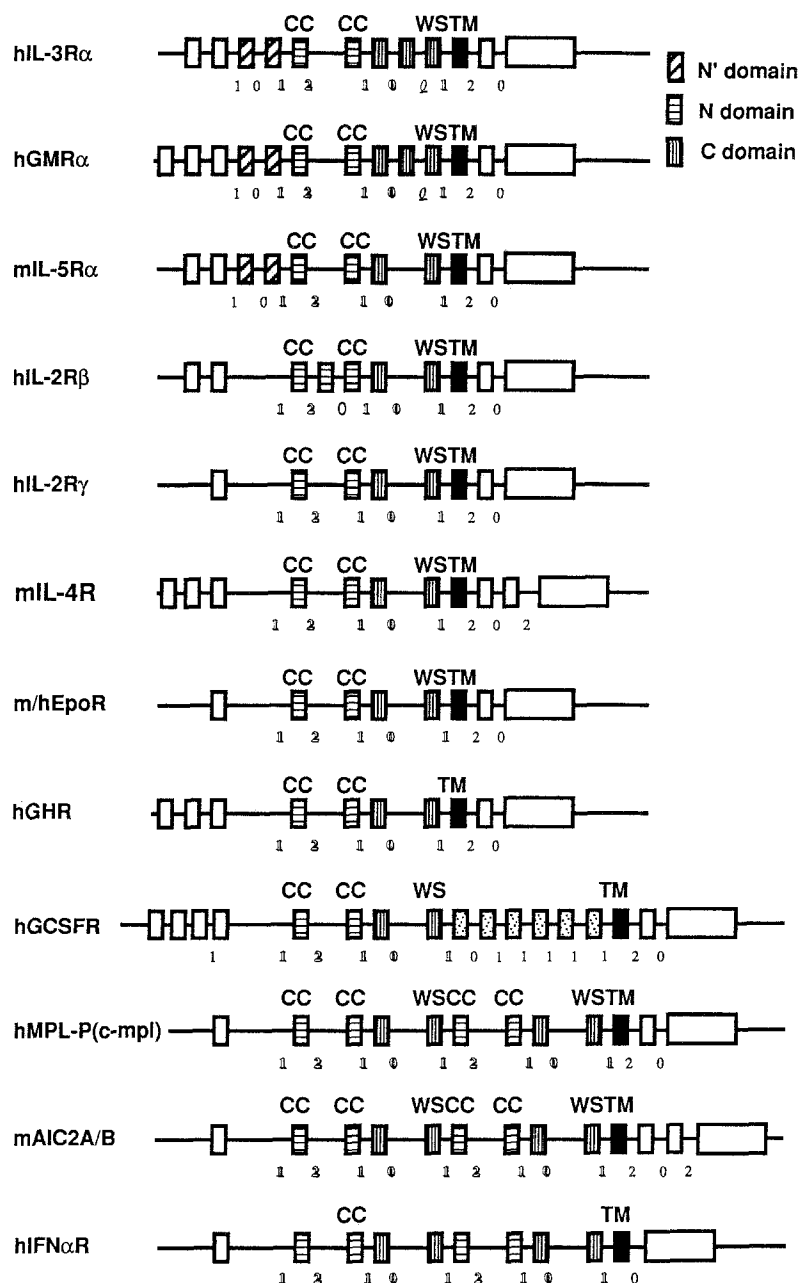
From the viewpoint of the gene structure, it is likely that cytoplasmic segments of type I CRSF also derived from a common ancestral gene. We have no idea what protein module this ancestral gene encoded. It is intriguing to acquire knowledge about the putative ancestral protein module of the cytoplasmic segment of type I CRSF, because it should give us significant information about the potency in sharing mechanisms of signal transduction. In this regard, it is noted that gene structures in the cytoplasmic segment of type I and type II CRSF members are different. On the contrary, those in the extracellular segment of type I and type II CRSF members are similar (Fig. 4, A).

## DISCUSSION

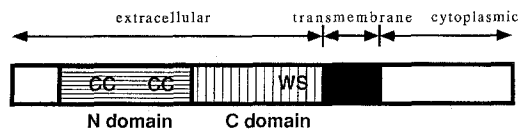
### CRSF members evolved from a putative common ancestral gene: A possible model for gene evolution of the CRSF

Bazan<sup>19</sup> originally classified cytokine receptors into classes 1 and 2. He also pointed out the

features of the protein structure of the extracellular segment of the class 1 and class 2 cytokine receptors and predicted their protein structure as a tandem array of Fn3-like domains; each domain contains seven  $\beta$  strands, which together form a sandwich of two antiparallel  $\beta$  sheets, one with four strands and one with three, with the same topology in each domain. And De Vos et al.<sup>23</sup> revealed that the three-dimensional structure of growth hormone receptor, in the first study on crystal structure in the CRSF members, is identical to topology of sandwich of domain D2 of CD4 and domain D2 of chaperone protein PapD, which shows the two-domain structure and the presence in each domain of two  $\beta$  sheets. Several studies on crystal structure of cytokines<sup>24, 25</sup> support the idea that CRSF members are likely to display protein structure similar to that of growth hormone receptor. Similarity in the expected protein structure, as well as in the genomic structure, among CRSF members indicates the related process of gene evolution in the manner in which exon insertion and exon duplication, two major mechanisms of exon shuffling, play an important role in their evolution.<sup>26-28</sup> Because we found the 1-2-1-0-1 rule

**A**

## protein structure of the CRSF



## gene structure of two Fn3-like domains of the CRSF

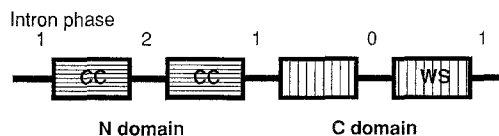
**B**

**FIG. 4.** Schematic representation of the gene structure of the CRSF. **A**, Comparison of gene structures of type I and II CRSF members available with intron phases. All extracellular segments follow the 1-2-1-0-1 rule, as shown here. An additional intron of phase 0 in the C domain is unique to IL-3R $\alpha$  and GMR $\alpha$  genes. Gene structures of the cytoplasmic segments also show similarity among type I CRSF members. See text for further details. Conserved cysteine residues and the WSxWS motifs are denoted by CC and WS, respectively. TM indicates transmembrane region. *hIL-2R $\beta$* , Human IL-2 receptor  $\beta$  subunit; *hIL-2R $\gamma$* , human IL-2 receptor  $\gamma$  subunit; *m/hEpoR*, mouse and/or human erythropoietin receptor; *hGH*, human growth hormone receptor; *hGCSFR*, human granulocyte colony-stimulating factor receptor; *hIFN $\alpha$ R*, human interferon- $\alpha$  receptor. **B**, The relation of protein structure and gene structure should explain why the sequential arrangement of intron phases within two Fn3-like domains as 1-2-1-0-1 can be the rule.

of the Fn3-like domains located in the extracellular segments of the type I and II CRSF members, we assume that extracellular segments of these cytokine receptors derive from a common ancestral gene. According to a model shown in Fig. 5, *B*, this gene segment was possibly formed from a prototype exon, encoding an Fn3-like module, by internal tandem duplication and intron insertion.

The derivation of the WSxWS motif in the C domain is also intriguing. To date, the role of the structural feature of the WSxWS motif in function is not clear enough. Several studies indicate that the WSxWS motif is important for interacting with other subunit(s).<sup>29</sup> Crystal analysis of human prolactin receptor (hPRLR) described the structural feature of the WSxWS box as part of a much larger, highly organized pattern covering much of the surface of the four-stranded sheet and consisting of alternating hydrophobic and charged groups. Sequence alignment of the CRSF shows analogous folding and arrangement in all class 1 receptors.<sup>30</sup> Although the first WSxWS sequences in the Fn3 module, for which the phylogenetic tree from bacteria to mammals has been studied,<sup>20</sup> is unknown to date, the WSxWS motif is found not only in CRSF but also in the complement precursor proteins C7, C8, and C9 and in thrombospondin and properdin. Fibronectin is known to bind to the specific sequences of WSxWS in the thrombospondin type I domain of thrombospondin.<sup>31, 32</sup> It was also found that UNC-5, a C. elegans receptor, which acts in migrating cells and pioneering neurons, possesses a motif of "DGG-WSXWSXWSXC" in the thrombospondin type I domain, which resembles the WSxWS motif in the C domain of the CRSF.<sup>33</sup> When the first WSxWS

acquired by an Fn3 module was identified, some linkage between the WSxWS motif of the CRSF and a motif found in the thrombospondin domain, which is similar to the WSxWS motif, could be established.

#### **A subfamily of IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$ can be distinguished from other type I CRSF members by the third Fn3-like domain**

Among the CRSF members, only three receptor subunits of IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  possess the third Fn3-like domains, N' domain,\* in their N' termini. No unique motif has been found in the N' domain thus far, which differs from the N' domain including four conserved cysteine residues and the C domain, including the WSxWS motif. The gene structure of the N' domain can be compared with the hIL-3R $\alpha$ ,<sup>5</sup> mIL-3R $\alpha$ ,<sup>34</sup> hGMR $\alpha$ ,<sup>4</sup> and mIL-5R $\alpha$  genes.<sup>35</sup> In all of them, the boundaries of this third Fn3-like domain are surrounded by phase 1 introns and interrupted by a phase 0 intron. As a result, the sequential arrangement of intron phases in this domain is conserved as 1-0-1. In the same context of the N' domain and the C domain, the N' domain might be acquired by exon shuffling. For this reason, not only in view of the protein structure, but also of gene evolution, it appears that IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  can be classified as a subfamily possessing the N' domain from other type I cytokine receptor subunits.

Among these three  $\alpha$ -subunit genes, we also found an intriguing difference in structure: the presence of the additional intron in the C domain, unique to the IL-3R $\alpha$  and GMR $\alpha$  genes. This additional intron is found in the hIL-3R $\alpha$  (intron 8), the mIL-3R $\alpha$  (intron 8), and the hGMR $\alpha$  (intron 9) genes, whereas it is not found in the mIL-5R $\alpha$  gene.

From these analyses, we choose to assume that the IL-3R $\alpha$  and GMR $\alpha$  genes evolved separately from the IL-5R $\alpha$  gene after these three  $\alpha$ -subunit genes emerged from a common ancestral gene acquiring the N' domain. Because most of the sequential arrangement of intron phases, known thus far in the CRSF displays a basic 1-2-1-0-1 arrangement, the IL-3R $\alpha$  and the GMR $\alpha$  genes differ most widely from them in gene evolution (Fig. 5, *A*). Several groups of subunits have a common subunit (i.e., gp130 shared by IL-6R, LIF-BP, CNTF, OSM, and IL-11R; IL-2R $\gamma_c$  shared by IL-2, 4, 7, 9, 13, and 15R) in the type I cytokine receptor subunits. Because much of the gene structure of these other receptor-

\*J. F. Bazan. Personal communication

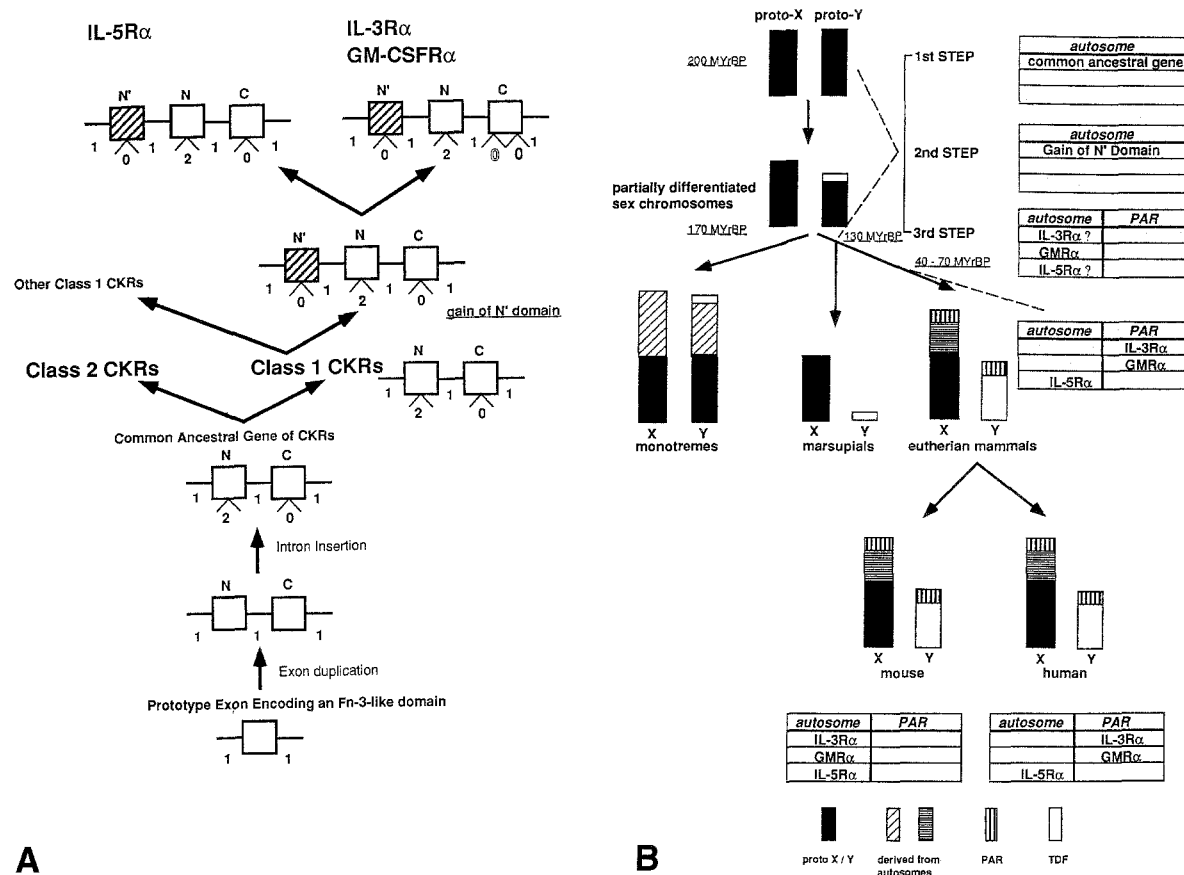


**FIG. 5.** Gene evolution of the CRSF. **A**, Model for gene evolution of the CRSF. A possible mechanism and process of CRSF diversification. Boxes and lines represent exons and introns, respectively. A common ancestral gene for the CRSF was generated from a prototype exon by internal duplication and intron insertion. The subfamily of IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  diversified from other type I cytokine receptors to acquire the N domain, and IL-3R $\alpha$ /GMR $\alpha$  evolved separately from IL-5R $\alpha$  with insertion of an additional intron into the C domain. **B**, Schematic representation of a putative process for the gene evolution and gene transfer of the subfamily of IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  in relation to the evolution of X chromosome in mammals. The GMR $\alpha$  gene emerged on an autosome at least before the divergence of marsupials and eutherian mammals, at least 130 MYrBP, early in the 200 million-year history of mammals. A putative common ancestral gene for IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  genes had emerged before that, possibly on an autosome. It is not clear, to date, whether genes of this subfamily had emerged before mammals diverged from avians or earlier. Because of Ohno's law of conservation of the X chromosome, recent studies on X inactivation prefer the idea that many genes on the pseudoautosomal region had been transferred from an autosome to the X chromosome first; and thereafter, transfer from the X chromosome to an autosome occurred in only mouse lineage after the divergence of human and mouse. TDF, Testis-determining factor gene; PAR, pseudoautosomal region; CKRs, cytokine receptors.

sharing systems is not available yet, we cannot know whether it is possible to deduce the presence of each common origin for each and every one of these groups. Anyway, we have not found any exceptions to our model for gene evolution of the CRSF thus far.

#### When did IL-3R $\alpha$ and GMR $\alpha$ genes emerge? Possible phylogeny of the CRSF

There is information about cytokines and cytokine receptors of species other than human beings

and mice. In addition to mammals, interferon (IFN) cDNA is cloned in fish<sup>36</sup> and chicken.<sup>37</sup> Sequence comparison reveals that chicken myelocytic growth factor (cMGF) is distantly related to granulocyte colony-stimulating factor (G-CSF) and IL-6 of mammals.<sup>38</sup> (Bazan<sup>39</sup> shows cMGF is closer to m/hG-CSF than to m/hIL-6 in the phylogenetic tree.) There is also the possibility that fish have an IL-2-like substance.<sup>40-42</sup> As for the CRSF, only the chicken prolactin and growth hormone

receptors are reported.<sup>43, 44</sup> From these studies, we can deduce that type I CRSF members at least existed in mammals and birds and that type II CRSF members, such as IFNR, should at least exist in fish.

Additionally, if an IL-2-like substance of fish is established as identical to mammalian IL-2, phylogeny of type I CRSF members can be traced from divergence of birds and mammals to that of vertebrates and invertebrates at least.

As for IL-3R and GMR, there are few studies of species other than human beings and mice. In human beings the IL-3R $\alpha$  (*IL3RA*) and the GMR $\alpha$  (*CSF2RA*) genes are located in the X-Y pseudoautosomal region,<sup>45, 46</sup> lying within 190 kb in the order of pter-*CSF2RA-IL3RA*-cen, whereas the mIL-3R $\alpha$  gene is located on chromosome 14 and the mGMR $\alpha$  gene, on chromosome 19 in mice.<sup>1</sup>

Several studies on evolution of the X chromosome have provided us with useful information about gene evolution of GMR $\alpha$ . Spencer et al.<sup>47</sup> and Graves et al.<sup>48</sup> showed that the locus of GMR $\alpha$  (*GMCSF*) would be on an autosome in marsupials. The pseudoautosomal region is supposed to have originally been autosomal, being added to the eutherian X after the marsupials' divergence.<sup>48</sup> Together with this information and our model for evolution of the IL-3R $\alpha$ , GMR $\alpha$ , and IL-5R $\alpha$  genes, we can show that the GMR $\alpha$  gene emerged in an autosome, at least before divergence of eutherian mammals and marsupials, about 130 million years ago (MYrBP), early in the 200 million-year history of mammals (Fig. 5, B). Evidence indicates that translocation of many genes from autosomes to X chromosome might have occurred in the 40 to 70 million years separating the metatherian and eutherian progenitors.<sup>49</sup>

Although there is no definitive evidence indicating that the GMR $\alpha$  gene transfers to the X chromosome before divergence in mice and human beings, it is likely that the GMR $\alpha$  gene was transferred from an autosome to the X chromosome after divergence in marsupials and eutherian mammals and then thereafter transferred to another autosome (chromosome 19 in mice) only in mouse lineage after branching off from human lineage.<sup>50</sup> Because there is no suggestion, thus far, about IL-3R $\alpha$  and IL-5R $\alpha$  in marsupials or monotremes, we cannot speculate anything about their emergence related to an evolution model of the X chromosome. Anyway, it is reasonable to assume that at least their putative common origin must have emerged on an autosome before GMR $\alpha$  had emerged on an autosome (Fig. 5, B).

### Promoter activity of the 5' upstream region of IL-3R $\alpha$ and GMR $\alpha$ genes

We determined the nucleotide sequence of the 5' upstream region of the hGMR $\alpha$  gene and found that it contains no typical TATA or Sp1 motifs, which are also seen in the hIL-2R $\gamma$  and human granulocyte-colony stimulating factor receptor (hGCSFR) genes. Around the transcription initiation sites of some "TATA-less," as well as TATA-containing promoters, there are specific sequences, called *initiator elements*, which contain the necessary information for basal transcription. At the major initiation site of the hGMR $\alpha$  gene, we found a sequence 1 (+1 -ACTCT- +5), which is homologous to the 3' half of the TdT-type initiator element. Therefore it is possible that upstream elements of the hGMR $\alpha$  gene, such as the purine-rich stretch at -30 (GGGGAGGGGAGGGG), function cooperatively with the initiator region and TATA-binding protein. Additionally, the hIL-2R $\gamma$  and the hGCSFR genes show a striking homology with the hGMR $\alpha$  gene around the transcription initiation sites (GGGAGGG). It suggests the possibility of a sharing mechanism for basal transcription of these three genes.

We also pointed out the potential binding sites for NF-IL6, ets family, and basic helix-loop-helix family transcription factors. Although the physiologic significance of these regulatory elements remains to be determined, we pointed out that the GAGGAA sequences at -48 to -43 and -5 to +1 would be noteworthy. This sequence contains the binding site for PU.1, an ets family transcription factor principally expressed in macrophages and B cells. After our report was published, it was reported that PU.1 at -45 plays an important role in the tissue-specific expression of the hGMR $\alpha$  gene in U937 cells.<sup>51</sup>

As for mIL-3R $\alpha$ , Miyajima et al.<sup>34</sup> analyzed the promoter activity of the 5' upstream region of the mIL-3R $\alpha$  gene. The mIL-3R $\alpha$  gene shows cell-specific expression. The 0.8 kb *PstI-StuI* fragment of the 5' upstream region of the mIL-3R $\alpha$  gene linked to the promoterless firefly luciferase gene directs the transcription in mouse mast cells (PT18) but not in mouse fibroblasts (NIH3T3) or in T cells (CTLL2).

### Conclusion

We have analyzed the structures of hIL-3R $\alpha$  and hGMR $\alpha$  genes. Comparison of gene structures, including those of hIL-3R $\alpha$  and hGMR $\alpha$  genes, with all of the CRSF genes available to date makes it possible to describe the common features of

them systematically. Through these analyses, we have proposed a model for gene evolution of CRSF members.

## REFERENCES

- Miyajima A, Mui AL-F, Ogorochi T, Sakamaki K. Receptors for granulocyte-macrophage colony-stimulating factor, interleukin-3, and interleukin-5. *Blood* 1993;82:1960-74.
- Kishimoto T, Taga T, Akira S. Cytokine signal transduction. *Cell* 1994;76:253-62.
- Miyajima A, Kitamura T, Harada N, Yokota T, Arai K. Cytokine receptors and signal transduction. *Annu Rev Immunol* 1992;10:295-331.
- Nakagawa Y, Kosugi H, Miyajima A, Arai K, Yokota T. Structure of the gene encoding the  $\alpha$  subunit of the human granulocyte-macrophage colony stimulating factor receptor: implications for the evolution of the cytokine receptor superfamily. *J Biol Chem* 1994;269:10905-12.
- Kosugi H, Nakagawa Y, Hotta T, et al. The structure of the gene encoding the  $\alpha$  subunit of the human interleukin 3 receptor. *Biochem Biophys Res Commun* 1995;208:360-7.
- Kitamura T, Sato N, Arai K, Miyajima A. Expression cloning of the human IL-3 receptor cDNA reveals a shared  $\beta$  subunit for human IL-3 and GM-CSF receptors. *Cell* 1991;66:1165-74.
- Gearing DP, King JA, Gough NM, Nicola NA. Expression cloning of a receptor for human granulocyte-macrophage colony-stimulating factor. *EMBO J* 1989;8:3667-76.
- Breathnach R, Chambon P. Organization and expression of eukaryotic split genes coding for proteins. *Annu Rev Biochem* 1981;50:349-83.
- Britten J, Baron WF, Stout DB, Davidson EH. Sources and evolution of human Alu repeated sequences. *Proc Natl Acad Sci USA* 1988;85:4770-4.
- Shibuya H, Yoneyama M, Nakamura Y, et al. The human interleukin-2 receptor  $\beta$  chain gene: genomic organization, promoter analysis and chromosomal assignment. *Nucleic Acids Res* 1990;18:3697-703.
- Noguchi M, Adelstein S, Cao X, Leonard WJ. Characterization of the human interleukin-2 receptor  $\gamma$  chain gene. *J Biol Chem* 1993;268:13601-8.
- Seto Y, Fukunaga R, Nagata S. Chromosomal gene organization of the human granulocyte colony-stimulating-factor receptor. *J Immunol* 1992;148:259-66.
- Ito Y, Seto Y, Brannan CI, et al. Structural analysis of the functional gene and pseudogene encoding the murine granulocyte colony-stimulating-factor receptor. *Eur J Biochem* 1994;220:881-91.
- Penny LA, Forget BG. Genomic organization of the human erythropoietin receptor gene. *Genomics* 1991;11:974-80.
- Gorman DM, Itoh N, Jenkins NA, Gilbert DJ, Copeland NG, Miyajima A. Chromosomal localization and organization of the murine genes encoding the  $\beta$  subunits (AIC2A and AIC2B) of the interleukin 3, granulocyte/macrophage colony-stimulating factor, and interleukin 5 receptors. *J Biol Chem* 1992;267:15842-8.
- Wrighton N, Campbell LM, Harada N, Miyajima A, Lee F. The murine interleukin-4 receptor gene: genomic structure, expression and potential for alternative splicing. *Growth Factors* 1992;6:103-18.
- Mignotte V, Vigon I, Boucher de Crevecoeur E, Romeo PH, Lemarchandel V, Chretien S. Structure and transcription of the human c-mpl gene (MPL). *Genomics* 1994;20:5-12.
- Sharp PA. Speculations on RNA splicing. *Cell* 1984;23:643-6.
- Bazan JF. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc Natl Acad Sci USA* 1990;87:6934-8.
- Bork P, Doolittle R. Proposed acquisition of an animal protein domain by bacteria. *Proc Natl Acad Sci USA* 1992;89:8990-4.
- Patthy L. Introns and exons. *Curr Opin Struct Biol* 1994;4:383-92.
- Murakami M, Narazaki M, Hibi M, et al. Critical cytoplasmic region of the interleukin 6 signal transducer gp 130 is conserved in the cytokine receptor family. *Proc Natl Acad Sci USA* 1991;88:11349-53.
- De Vos AM, Ultsch M, Kossiakoff AA. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 1992;255:306-12.
- Robinson RC, Grey LM, Staunton D, et al. The crystal structure and biological function of leukemia inhibitory factor: implications for receptor binding. *Cell* 1994;77:1101-16.
- Milburn MV, Hassel AM, Lambert MH, et al. A novel dimer configuration revealed by the crystal structure at 2.4 angstrom resolution of human interleukin-5. *Nature* 1993;363:172-6.
- Gilbert W. Why genes in pieces? *Nature* 1978;271:501.
- Patthy L. Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett* 1987;214:1-7.
- Mattick J. Introns: evolution and function. *Curr Opin Gene Dev* 1994;4:823-31.
- Yoshimura Y, Zimmers T, Neumann D, Longmore G, Yoshimura Y, Lodish HF. Mutations in the Trp-Ser-X-Trp-Ser motif of the erythropoietin receptor abolish processing, ligand binding, and activation of the receptor. *J Biol Chem* 1992;267:11619-25.
- Somers W, Ultsch M, De Vos AM, Kossiakoff AA. The x-ray structure of a growth hormone-prolactin receptor complex. *Nature* 1994;372:478-81.
- Sipes JM, Guo N, Negre E, Vogel T, Krutzsch HC, Roberts DD. Inhibition of fibronectin binding and fibronectin-mediated cell adhesion to collagen by a peptide from the second type I repeat of thrombospondin. *J Cell Biol* 1993;121:469-77.
- Vogel T, Guo NH, Krutzsch HC, et al. Modulation of endothelial cell proliferation, adhesion, and motility by recombinant heparin-binding domain and synthetic peptides from the type I repeats of thrombospondin. *J Cell Biochem* 1993;53:74-84.
- Leung-Hagesteijn C, Spence AM, Stern BD, et al. UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* 1992;71:289-99.
- Miyajima I, Levitt L, Hara T, et al. The murine interleukin-3 receptor  $\alpha$  subunit gene: chromosomal localization, genomic structure and promoter function. *Blood* 1995 (in press).
- Imamura F, Takaki S, Akagi K, et al. DNA Cell Biol 1994;13:283-92.
- Tamai T, Shirahata S, Noguchi T, Sato N, Kimura S, Murakami H. Cloning and expression of flatfish (*Paralichthys olivaceus*) interferon cDNA. *Biochim Biophys Acta* 1993;1174:182-6.

37. Sekellick MJ, Ferrandino AF, Hopkins DA, Marcus PI. Chicken interferon gene: cloning, expression, and analysis. *J Interferon Res* 1994;14:71-9.
38. Leutz A, Damm K, Sterneck E, et al. Molecular cloning of the chicken myelomonocytic growth factor (cMGF) reveals relationship to interleukin 6 and granulocyte colony stimulating factor. *EMBO J*. 1989;8:175-81.
39. Bazan JF. Neuropoietic cytokines in the hematopoietic fold. *Neuron* 1991;7:197-208.
40. Eitan S, Zisling R, Cohen A, et al. Identification of an interleukin 2-like substance as a factor cytotoxic to oligodendrocytes and associated with central nervous system regeneration. *Proc Natl Acad Sci USA* 1992;89:5442-6.
41. Eitan S, Schwartz M. A transglutaminase that converts interleukin-2 into a factor cytotoxic to oligodendrocytes. *Science* 1993;261:106-8.
42. Eitan S, Solomon A, Lavie V, et al. Recovery of visual response of injured adult rat optic nerves treated with transglutaminase. *Science* 1994;264:1764-8.
43. Tanaka M, Maeda K, Okubo T, Nakashima K. Double antenna structure of chicken prolactin receptor deduced from the cDNA sequence. *Biochem Biophys Res Commun* 1992;188:490-6.
44. Duriez B, Sobrier ML, Duquesnoy P, et al. A naturally occurring growth hormone receptor mutation: in vivo and in vitro evidence for the functional importance of the WS motif common to all members of the cytokine receptor superfamily. *Mol Endocrinol* 1993;7:806-14.
45. Milatovich A, Kitamura T, Miyajima A, Francke U. Gene for the alpha subunit of the human interleukin 3 receptor (*IL3RA*) localized to the X-Y pseudoautosomal region. *Am J Human Genet* 1993;53:1146-53.
46. Kremer E, Baker E, D'Andrea RJ, et al. A cytokine receptor gene cluster in the X-Y pseudoautosomal region? *Blood* 1993;82:22.
47. Spencer JA, Sinclair AH, Watson JM, Graves JAM. Genes on the short arm of the human X chromosome are not shared with marsupial X. *Genomics* 1991;11:339-45.
48. Graves JAM, Schmidt MM. Mammalian sex chromosomes: design or accident? *Curr Opin Gene Dev* 1992;2:890-901.
49. Mitchell MJ, Woods DR, Wilcox SA, Graves JAM, Bishop CE. The marsupial Y chromosome encodes a homologue of the mouse Y-linked candidate Spermatogenesis gene *Ube1y*. *Nature* 1992;359:528-31.
50. Lyon MF. Evolution of the X chromosome. *Nature* 1990;348:585-6.
51. Hohaus S, Tennen DG. The transcription factor PU.1 regulates expression of the granulocyte-macrophage CSF receptor alpha promoter. *Blood* 1994;84(suppl 1):150a (2024).
52. Ashworth A, Kraft A. Cloning of a potentially soluble receptor for human GM-CSF. *Nucleic Acids Res* 1990;18:7178.
53. Raines MA, Liu L, Quan SG, et al. Identification and molecular cloning of a soluble human granulocyte-macrophage colony-stimulating factor receptor. *Proc Natl Acad Sci USA* 1991;88:8203-7.
54. Crosier KE, Wong GG, Mathey-Prevot B, Nathan DG, Sieff CA. A functional isoform of the human granulocyte/macrophage colony-stimulating factor receptor has an unusual cytoplasmic domain. *Proc Natl Acad Sci USA* 1991;88:7744-8.