Volume 14 Number 20 1986 Nucleic Acids Research

Genomic sequence for human prointerleukin 1 beta: possible evolution from a reverse transcribed prointerleukin 1 alpha gene

Burton D.Clark[1], Kathleen L.Collins[2], Melinda S.Gandy[1], Andrew C.Webb[2] and Philip E.Auron[1,3]

[1]Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, MA 02139, [2]Department of Biological Sciences, Wellesley College, Wellesley, MA 02181 and [3]Department of Medicine, The New England Medical Center-Tufts University School of Medicine, Boston, MA 02111, USA

Received 10 July 1986; Revised and Accepted 16 September 1986

ABSTRACT

We have isolated the human prointerleukin 1 (proIL-1) beta gene from leukocyte and fetal liver libraries. The nucleotide sequence and its gene organization reveals that the proIL-1β gene is composed of seven exons with a primary transcription product length of 7,008 nucleotides. The exon sequence agrees well with that of the human proIL-1β cDNA. Features of interest within the transcriptional unit include conventionally positioned **TATA**, **CAT**, and poly-adenylation signals for gene regulation, as well as the signatures of gene duplication via retrotransposition in the form of flanking direct repeats and a genomic poly A tail. The genomic organization of the proIL-1β gene with respect to the number and position of exon boundaries is strikingly similar to that of the recently reported human proIL-1α gene. Therefore, we hypothesize that the proIL-1β may have arisen by a reverse transcriptase mediated duplication of the related alpha gene.

INTRODUCTION

The isolation of large quantities of pure interleukin 1 (IL-1) proteins required for the identification of biological properties and activities has been a problem for many years. One direction taken to solving this has been the cloning and sequencing of IL-1 cDNA from three different species. Analysis of these cDNA clones have identified two distinct forms of IL-1 precursor denoted proIL-1α and proIL-1β [1, 2, 3, 4]. Characterization and expression of the proIL-1α and proIL-1β cDNA has allowed extensive studies on the biological properties of IL-1 [5, 6]. However, little is known about the molecular regulation of IL-1 expression. As a first step in understanding the regulation, we report here the cloning, sequencing, and characterization of the entire human proIL-1β gene and compare it to the recently published human proIL-1α gene sequence [7].

The existence of two distantly related proIL-1 genes coding for two functionally similar proteins raises evolutionary issues of origin and diversification. Of particular interest, the proIL-1β gene may be derived from the proIL-1α gene via an RNA-mediated duplication-transposition event. Within eukaryotic genomes, there are many examples of processed genes possibly formed by the reverse transcription of cellular RNA. These

genes may form in the germline by the activity of cellular reverse transcriptase with the final cDNA transcript intergrating into the host chromosome. These processed genes, or retroposons [8], are usually lacking introns, bounded by direct repeats, and have the remnants of a 3' poly A tail. Examples of human retroposons include genomic DNA related to β-tubulin, immunoglobin $C_\epsilon 3$, and metallothionein II [9, 10, 11, 12]. These sequences, also described as pseudogenes, are not functional due to loss of regulatory elements and/or introns. Recently, one of the two rat insulin genes, preproinsulin I, has been reported to be a functional retroposon [13]. This observation supports an earlier suggestion that the generation of retrotransposed pseudogenes serves an evolutionary role in gene duplication, enabling a mechanism for the transposition of a gene copy to a distant site [14]. In this study, we report a similar observation for the human proIL-1β gene which is flanked by direct repeats and has remnants of a 3' poly A tail comparable to rat preproinsulin I.

In addition, given the low degree of sequence homology observed between the proIL-1 alpha and beta cDNAs [2], an unexpected finding was that the two genes contain exons similar in both length and the positioning of the intron/exon boundaries. These boundaries segregate the IL-1 protein domains reported to be important for biological activity [15, 5] into separate exons. Such conservation of gene structure and organization is consistent with both a common ancestry as well as a functional constraint on evolutionary diversification.

MATERIALS AND METHODS

Materials

Restriction endonucleases were from New England Biolabs, International Biotechnologies, Inc. or Boehringer Mannheim. Nick translation employed a kit from Bethesda Research Laboratories. The SP6 RNA polymerase and plasmid pSP64 were obtained from Promega Biotec, Inc. The radiochemicals α^{32}P-UTP (3000 Ci/mmol) and α^{32}P-dCTP (3000 Ci/mmol) were purchased from Amersham Corp. and α^{35}S-dATP (500 Ci/mmol) was purchased from DuPont-New England Nuclear. M13 deletion clones were prepared using the Cyclone system from International Biotechnologies, Inc. Human leukocyte DNA was a kind gift from Clontech Inc., Palo Alto, CA. and human placenta DNA was purchased from Sigma Chemical Co.

Southern blot hybridization

Human DNA (10μg), isolated by the method of Farmer *et al.*, was digested with restriction endonucleases and resolved by electrophoresis on 0.8 % agarose gels. The DNA fragments were then electroblotted to Gene Screen membrane (DuPont-New England Nuclear) and hybridized using conditions described by the manufacturer. A

Pst I - *Xho* I fragment containing the human proIL-1β cDNA from pcD1218 [1] was nick translated and used as a probe.

Cloning, sequencing, and blotting techniques

A bacteriophage lambda Charon 4A library containing inserts of human fetal liver DNA that were partially digested with *Alu* I - *Hae* III was kindly provided by T. Maniatis [16]. A lambda EMBL 3 library containing *Mbo* I inserts of partially digested human leukocyte DNA was purchased from Clontech Inc., Palo Alto, CA. Both libraries were screened by the procedure of Benton and Davis [17] using either a nick translated 1075 base pair *Fnu*D II - *Xmn* I DNA fragment of the proIL-1β cDNA [1] as hybridization probe or a 1350 base RNA riboprobe transcribed from the proIL-1β cDNA cloned into the plasmid pSP64 [18]. Hybridization positive phage clones were purified and then amplified to isolate recombinant phage DNA. Clones containing proIL-1β genomic DNA were characterized by restriction endonuclease mapping and fragments containing exons identified by Southern blot hybridization with proIL-1β cDNA as probe. DNA fragments of phage clones containing exons were isolated by electroelution and ligated into one of the M13 vectors, mp8, mp9, or mp11 and mp19. A deletion series of the M13 clones was prepared by the method of Dale *et al.* [19] before being sequenced by the method of Sanger *et al.* [20].

Computer techniques

Sequence assembly utilized a new DNA sequence assembly system written by Dr. William A. Gilbert of the Massachusetts Institute of Technology Whitaker College Computer Facility (personal communication). This program, called Multiple Sequence Editor (MSE), provides a rapid interactive method of assembling DNA sequence data. Plots of positional base preference and splice acceptor and donor sites were generated using the software described by Roger Staden [21, 22, 23]. Other consensus sequence searches used utilities such as FitConsensus, Consensus, and WordSearch which are contained within the University of Wisconsin Genetics Computer Group (UWGCG) software package [24].

RESULTS AND DISCUSSION

Isolation of genomic clones

Two different human genomic libraries were used in our isolation of the proIL-1β genomic sequences. The first was a human fetal liver library carried in the λCharon 4A vector [16] and the other a human leukocyte library carried in the λEMBL 3 vector. A total of 1.8 x 10^6 bacteriophage plaques of the fetal liver genomic library and 2.4 x 10^6 plaques of the leukocyte genomic library were screened using hybridization probes and conditions as described in Materials and Methods. Such screening resulted in the isolation of 15 hybridization positive clones from the fetal liver genomic library and 12 clones from the leukocyte genomic library.

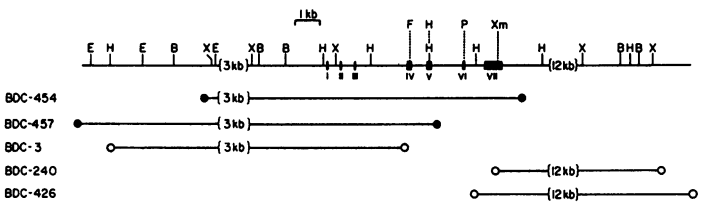


Figure 1. Schematic representation of the organization of the human proIL-1β gene locus. Solid boxes in the upper line represent the location of exon sequences within the approximately 38 kb of sequence contained within the overlapping clones indicated below. Clones derived from the λ Charon 4A library are shown bounded by open circles, whereas those derived from the λ EMBL 3 library are bounded by closed circles. Also shown are the locations of some restriction enzyme sites which are unique within the genomic sequence (solid vertical lines) as well as some sites which are unique within the cDNA sequence (broken vertical lines). The abbreviations used are: E=*Eco*R I; H=*Hind* III; B=*Bam*H I; X=*Xba* I; F=*Fnu*D II; P=*Pvu* II;and Xm=*Xmn* I. The 3 and 12 kb notation enclosed within braces indicate the location of regions of the sequence which do not contain these unique restriction enzyme sites.

As a result of initial characterization by restriction mapping and Southern blot analyses using proIL-1β cDNA as a probe, 5 unique clones were identified and used for detailed mapping and DNA sequencing. These clones, which share common restriction fragments, are shown schematically in Figure 1. The two different vectors used in this screening utilized two different sites for insertion of the genomic DNA fragments. The location of these sites (*Eco*R I for Charon 4A and *Bam*H I for EMBL 3) are indicated. One clone, designated as BDC-454, contains approximately 14 kb of genomic insert and represents the entire transcriptional unit of the proIL-1β gene. The gene consists of seven exons that range in size from 52 to 822 nucleotides and are interspersed by six introns of between 460 and 1981 nucleotides in length. Two clones, designated as BDC-457 and BDC-3 contain the first five and three exons, respectively, and extend the upstream sequence for the gene to approximately 12 kb beyond the transcriptional initiation site. Two other clones, designated BDC-240 and BDC-426 contain part or all of exon 7 and extend the downstream sequence to approximately 18 kb beyond the poly A addition site. The restriction mapping data derived from the overlapping clones isolated from the two different libraries (representing two distinct differentiated cell types) are identical, with the exception of the restriction sites derived from cloning. This renders unlikely any significant sequence reorganization due either to cellular differentiation or cloning. The absence of cloning artifacts is further supported by the similarity of genomic DNA digests to the restriction maps derived from the clones (Figure 2).

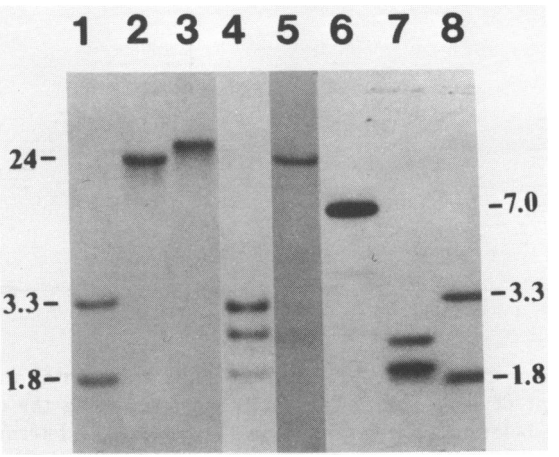


Figure 2. Autoradiogram of restriction fragments which hybridize to radio-labeled proIL-1β cDNA. Samples of 10 μg human monocyte (lanes 1-5) and 5 ng bacteriophage clone BDC-454 (lanes 6-8) DNA were digested with the following restriction endonucleases: (1) *Pvu* II, (2) *Bam*H I, (3) *Eco*R I, (4) *Hin*d III, (5) *Xba* I, (6) *Xba* I + *Sal* I, (7) *Hin*d III + *Sal* I, (8) *Pvu* II + *Sal* I. The sizes of selected fragments are indicated in kilobase pairs.

Comparison of total cellular and cloned DNA

Human monocyte and bacteriophage clone BDC-454 DNA were prepared and analyzed by the method of Southern [25] using ^{32}P-labeled proIL-1β cDNA [1] as a hybridization probe. Figure 2 shows the comparison between these DNAs digested with several different restriction endonucleases. A relatively simple banding pattern is seen suggesting that the proIL-1β gene is uniquely represented in the human genome. This observation confirms previous work from our lab which localized a single proIL-1β gene to the long arm of chromosome 2 at 2q13-2q21 [26]. The gene is located on a single 24 kb *Bam*H I or a 32 kb *Eco*R I fragment. The 1.8 kb bands for each DNA digested with *Hin*d III and *Pvu* II are doublets. The genomic DNA insert in clone BDC-454 can be excised as a single *Sal* I fragment. DNA from human monocytes and clone BDC-454 digested with *Pvu* II contain a common set of restriction fragments (lanes 1 and 8). As seen in Figure 2, differences exist between the sizes of restriction fragments derived from clone BDC-454 and total monocyte DNA. These differences are consistent with changes resulting from the techniques used for library construction which trim adjacent regions from the inserted genomic fragments. Therefore, since BDC-454 contains only a portion of the sequences flanking the proIL-1 gene, its banding pattern should be similar but not identical to that derived from total cellular DNA.

Sequence of genomic clones

Figure 3 shows a summary of the DNA sequencing strategy and the relative orientation

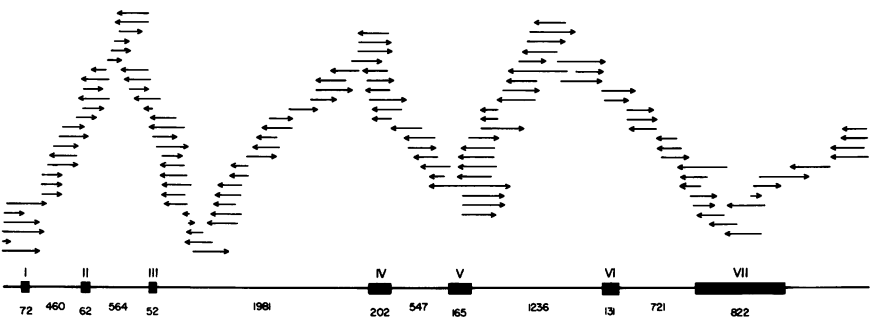


Figure 3. Sequencing strategy for the proIL-1β gene. Horizontal arrows represent the direction and extent of sequencing of the M13 subclones using the dideoxy terminator method. The line below the arrows represents the sequence determined and the bold regions (I through VII) locate exons. Numbers below the line indicate the length in base pairs of each region.

and sizes of the proIL-1β gene introns and exons. The entire nucleotide sequence is shown in Figure 4. The sequences of the overlapping clones isolated from the fetal liver and leukocyte libraries are identical. The predicted length of the primary transcription product, exclusive of polyadenylation, is 7,008 nucleotides. The exon sequences agree well with the previously published proIL-1β cDNA sequence [1] (vertical bars in Figure 4), however there are a total of 16 differences which are indicated by the absence of a vertical bar in Figure 4. Two of these differences (positions 7005 and 7008) are probably the result of trimming and substitution during post-transcriptional poly A addition. Only one of these nucleotide changes is located within the coding region and affects the nature of the proIL-1β translation product. This is replacement of a G for an A residue resulting in a lysine to glutamate substitution at amino acid position 6 of proIL-1β. This substitution was also reported by March *et al* following re-cloning and sequencing of the proIL-1β cDNA [2]. It is interesting that all the other substitutions reside in noncoding regions. However, the fact that 7 substitutions are contained within the first 100 nucleotides of the cDNA sequence suggests the possibility that reverse transcription errors during cDNA synthesis may be responsible for these upstream differences. Four of the six differences in the 5' untranslated region were previously reported by March *et al.* [2]. Additionally, these investigators reported 2 substitutions within the 3' untranslated region which do not correspond to a discordancy between our cDNA and genomic sequences (a C substitution for a G at position 6442 and a C insertion at position 6620 in Figure 4). The March *et al.* sequence agrees with our cDNA at position 6616 whereas the chromosomal sequence reported here contains an additional C residue. Since the March *et al.* data within the untranslated regions of the proIL-1β cDNA were truncated, we lack an additional sequence to compare the data at locations before position 11 or

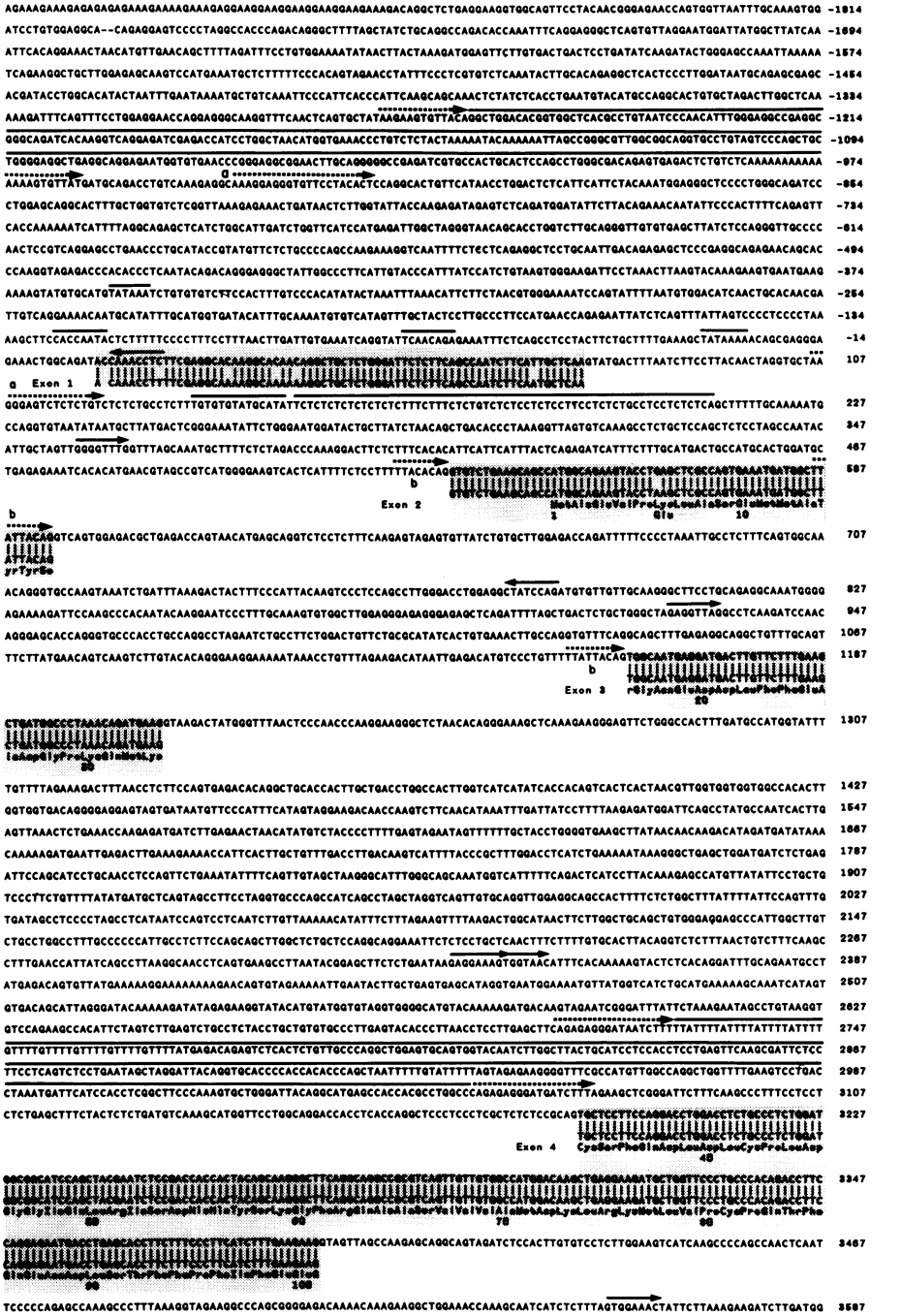
beyond 6643. Our earlier analysis of the transcriptional initiation site for this gene using primer extension revealed the possibility that our proIL-1β cDNA clone was one nucleotide short [27]. We therefore have positioned the transcriptional initiation site as indicated in Figure 4 at the appropriate adenine residue within the genomic sequence. This adjustment of the 5' end of the transcriptional initiation site corresponds to the additional C residue within the genomic sequence in comparison to the cDNA (Figure 4). We conclude that the first 7 upstream substitutions contained within the genomic sequence are an accurate representation of the 5' region of the proIL-1β gene. However, the 7 downstream, non-poly A related substitutions, cannot yet be ascribed to polymorphism or sequencing error until further data in this region are obtained.

Genomic sequence data analysis

Location of potential regulatory sequences. The human proIL-1β genomic sequence shown in Figure 4 was subjected to data analysis using several computer implemented data reduction techniques (see Materials and Methods). Consensus sequence searches for transcriptional promoter motifs have revealed the location of a **TATA** box sequence (TATAAAA) [28] at position -31 and **CAT** boxes (TCAAC and CCAAT) [28, 29] at positions -75 and -126, respectively. The -126 sequence is totaly consistent with the consensus (CCAAT) unlike that at position -75, which is the usual location for these elements. However, a functional **CAT** box with the sequence CCAAC, which differs from the consensus, has been identified at position -68 for the human beta-like delta globin gene [30]. This suggests that the human proIL-1β **CAT** box at position -75 may also be functional. In comparison, the proIL1-α gene, which is transcribed at approximately one fiftieth the level of proIL-1β in human monocytes [2], has a poor **TATA** box consensus sequence (TACAAA) and no obvious **CAT** box.

Several potential viral enhancer-like core sequences [31] have also been located in both orientations within the sequence. These are found exclusively within introns, with one exception; namely, within the first ten nucleotides of the 5' untranslated region of exon 1. Another sequence located between positions 5313 and 5320 bears strong homology to the enhancer core consensus and is immediately adjacent to a sequence (positions 5295 to 5313) which is very similar (75% homologous) to the glucocorticoid regulatory element described by Karin *et al.* [32]. These sequences are located in intron 5 which contains other enhancer-like sequences, an SP 1-like transcription factor binding site between positions 4564 and 4573 [33], and the only short-range dyad symmetry element within the proIL-1β genome (positions 4272-4287). Interestingly, multiple copies of similar types of sequence motifs have been located within intron 6 of the human proIL-1α genome.

```
ACCCGGTACACTGTGTCC-T  HMT-IIA Glucocorticoid Receptor Element (GRE)
|||| | |  ||||||||  |
ACCCTGCATGCTGTGTCCCT  proIL-1 Beta
```



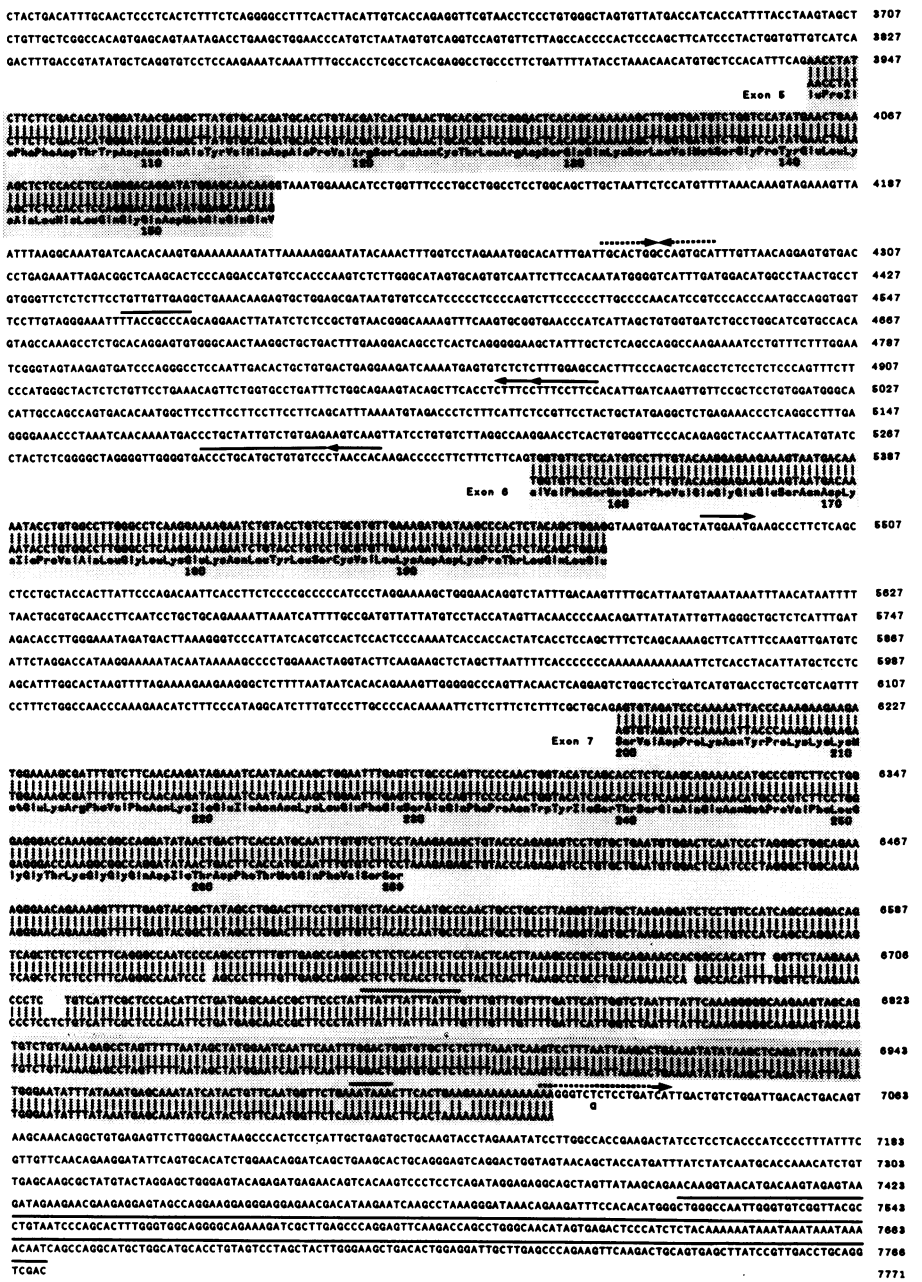


Figure 4. Nucleotide sequence of the human proIL-1β gene and comparison to the cDNA sequence. Key: overbars - sequences of interest described in text; solid arrows - potential enhancer cores; dotted arrows - direct and indirect repeats.

Intron 1 of the proIL-1β genome contains a Z-DNA-like alternating purine-pyrimidine tract (positions 135-148) immediately preceding a homo-pyrimidine sequence at positions 150-211. Such sequences have been identified as being essential for expression of a murine class II major histocompatibility gene which is also a macrophage cell product [34]. The highly conserved site required for polyadenylation [35] is located at positions 6992-6997. The conserved sequence YGTGTTYY, located about 30 nucleotides downstream from the poly A site in 67 of 100 mammalian genes studied [36], is absent in both proIL-1 genes.

Intron 3 contains a sequence representative of *Alu* repetitive DNA, possessing the usual flanking direct repeats and poly A tails (shown as poly T in Figure 4). Related sequences are also found flanking the entire IL-1β gene.

Is proIL-1 beta derived from proIL-1 alpha by retrotransposition? Examination of the genomic sequence corresponding to the 3' end of the cDNA reveals that a string of 14 A residues, interrupted by a single G residue registers exactly with the poly A tail found within the cDNA. Interestingly, a region of 17 nucleotides of the genomic sequence extending downstream from the genomic oligo A string is imperfectly repeated upstream both within intron 1 at position 106 as well as at position -940 (labeled as direct repeats **a** in Figure 4). The corresponding alignments of the two upstream sequences with the region adjacent to the oligo A tail are:

```
          106  AA---GGGAGTCT-CTC-TG---T  Intron 1
               ||   |||  ||| ||| ||   |
7009 AAAAAAAAAAA---GGG--TCT-CTCCTG-A-T  Downstream Oligo A Tail
             |||   |||  | |  ||||  | |
        -940 AAAGGAGGG--TGT--TCCTACACT  Upstream
```

These alignments reveal sufficient homology to suggest the possibility that the human proIL-1β gene is delineated by upstream and downstream direct repeats. The presence of a genomic oligo A tract that precisely aligns with the poly A tail contained within the mRNA transcript and positioned immediately adjacent to a direct repeat, has previously been considered to represent a signature of gene duplication derived from a reverse transcript transposed to a site distant from the original chromosomal location [14, 13]. The possibility that the oligo A tail within the proIL-1β gene corresponds to an A-rich region which fortuitously was used as a primer during cDNA cloning is unlikely since a poly A addition signal exists 13 nucleotides upstream of the poly A tail, and the next such site is located 654 nucleotides downstream. Furthermore, nucleotide length determination on Northern blots of IL-1β mRNA derived from various cell types [37] show very good agreement with the length of sequence derived from the cDNA. The absence of longer transcripts on Northern blots corresponding to the location of the

next poly A addition site argues against artifactual initiation at this location during molecular cloning.

Consequently, genomic organization of proIL-1β resembles *Alu* family repeats as well as some classes of pseudogenes [14, 38]. Interestingly, the recent discovery of an active retroposon corresponding to one of the two different rodent preproinsulin genes (types I and II) [13] establishes a precedent for what we believe is a similar situation for the relationship between the two different proIL-1 genes (alpha and beta). The model proposed for the generation of an active preproinsulin I retrotransposon involves transcription initiation of the preproinsulin II gene on mouse chromosome 7 at a cryptic upstream promoter, thus generating a transcript containing an RNA copy of the type II gene promoter. Subsequent polyadenylation, partial processing, and reverse transcription would yield an *in vivo* cDNA copy, which could then integrate into a staggered break on chromosome 6 generating flanking direct repeats upon DNA repair. Unlike retrotransposed inactive pseudogenes, which are derived from processed transcripts after normal initiation, preproinsulin I is an active gene because the aberrant long transcript carries an efficient promoter derived from upstream initiation.

The sequence of the human proIL-1β gene, like that of preproinsulin I, reveals the potential for upstream initiation at cryptic upstream promoter sequences. Two such promoter sequences are observed in the IL-1α sequence [7] and correspond well to similar sequences in IL-1β. Interestingly, a third distal upstream promoter in IL-1α is absent in the IL-1β sequence and corresponds in location to the direct repeat at position -940. Since some intergene homology also exists along the sequence extending from this site to the bonafide transcription start point, the derivation of the parental IL-1β transcript may have originated with initiation at the most distal alpha promoter. A striking difference between the preproinsulin I gene and proIL-1β is the absence of any intron processing. Since intron removal is commonly subsequent to polyadenylation [39], there is no *a priori* reason to believe that reverse transcription of an unprocessed primary transcript could not occur. The existence of a second direct repeat within intron 1 of the proIL-1β gene suggests an alternate mechanism for retrotransposition in which an incorrectly processed alpha transcript may have been the source of the parental genetic material. In this scheme the truncated beta retrotransposon could have been activated by insertion into a different gene downstream from a functonal promoter. However, the existence of upstream homology, potential promoters, and a direct repeat lends support to the first of these two models.

The presence of flanking direct repeats and an oligo A tract corresponding to the location of mRNA polyadenylation suggests the possibility of retrotransposition. However, gene duplication by a DNA-mediated event cannot be entirely eliminated. Retrotransposition is thought to generate gene copies which are usually distantly located

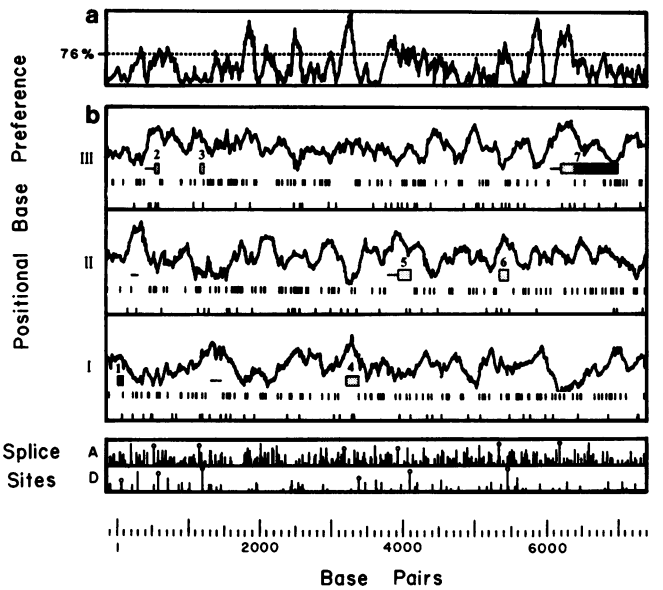


Figure 5. Translation and reading frame analysis for the proIL-1β gene. (a) Reading frame-independent analysis of uneven positional base preference as described by Staden [20] reveals the location of potential sites of protein coding-like sequences. The 76 % indicates the signal level for which there is 76 % agreement with the EMBL sequence database. (b) Reading frame-dependent analysis of positional base preference. Reading frames are labeled I, II, and III for the sequence initiating with nucleotide residues -133, -132, and -131, respectively. Shown below each trace are the numbered exons (1-7) indicating untranslated (solid boxes) and translated regions (gray-tone boxes). Also shown are the locations of some potential open reading frames (ORFs) within the gene (horizontal lines). The short vertical bars immediately below the exon boxes locate translational termination signals and the bars at the bottom of each panel indicate potential translation initiation sites. The locations of potential splice acceptors **A** and donors **D** are shown beneath the positional base preference plots as a histogram revealing relative agreement with the splice consensus sequence. Splice acceptors and donors utilized by human monocytes to process the proIL-1β transcripts are indicated by small open circles at each appropriate location within the histogram plot.

from the parental gene, while DNA-mediated events can generate either tandem duplication or distant translocation depending upon the nature of flanking repeated sequences [40]. For example, the presence of flanking *Alu* sequences of similar orientation in the proIL-1β gene suggests the possibility of a tandem gene duplication which could have been released as a circle for translocation to another site. However, without more detailed knowledge about the flanking sequence environment, and the chromosomal assignment for the IL-1α gene, a firm conclusion cannot be reached.

For several reasons we do not believe that the gene which we have cloned and

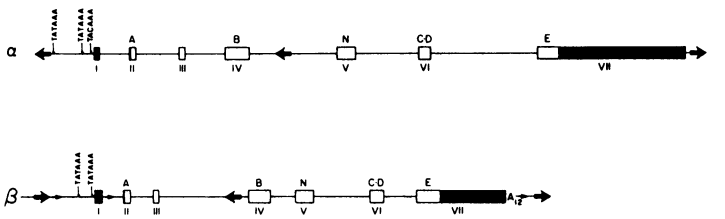


Figure 6. Comparison of the human proIL-1α and proIL-1β genes. The two genes are aligned schematically to reveal the similar exon organization. The relative lengths of the exons in both the alpha and beta sequences are to scale. This also holds true for the intron sequences, however the relative intron-exon lengths have been distorted in order to simplify the presentation. Open and closed boxes show the relative location of coding and noncoding regions within the exons which are numbered (roman numerals). Protein homology domains are indicated by the lettered notations **A** through **E**. The notation **N** indicates the exon containing the amino terminus of the processed 17,000 molecular weight IL-1β peptide. The large arrows locate the position and transcriptional orientation of *Alu* family sequences found within the IL-1 gene loci. The small arrows indicate the location of short direct repeats found in the proIL-1β gene adjacent to the oligo A tail (A_{12}) and upstream both within intron 1 as well as at position -940. The location of upstream promoter-like sequences are indicated (TATAAA or TACAAA).

sequenced is a pseudogene. First, there are no obvious defects within the coding region. Second, the sequence contiguous to splice sites agrees well with the Mount consensus for splice acceptors and donors [41] (see Figure 5) and corresponds perfectly with the "GT/AG rule" [42]. Furthermore, each splice acceptor, without exception, contains upstream lariat branch site consensus sequences [43, 44]. Finally, the presence of acceptable promoter sequences at the required upstream locations, and the absence of any other hybridizable signals either within the genomic libraries or on Southerns of total cellular DNA, strongly suggests that our sequence represents the active gene.

The structural organization of the two genes reveals a striking relationship which is summarized schematically in Figure 6. The organization of the exons of the two genes is quite similar both in length as well as in the location of the splice sites with respect to the protein sequences. The relationship between the splice sites and the protein sequences for the alpha and beta forms is shown in detail in Table I. In each instance, the splice site splits the coding region at locations which delineate the putative functional domains within the protein sequences [15, 5]. Amazingly each corresponding splice cuts the coding region of either alpha or beta proIL-1 within one amino acid of the aligned regions previously reported [15]. It is striking that two such distantly related molecules (i.e., 41% cDNA homology) should be so highly constrained not only in terms of the number of exons, but also in the precise register of exon boundaries with respect to protein domains. This conservation may simply reflect the functional importance of

Table I. Comparison of proIL-1 alpha and beta exon boundaries.

	proIL-1 Alpha		proIL-1 Beta	
Exon #	Inclusive Amino Acids	Protein Homology Domains	Inclusive Amino Acids	Protein Homology Domains
2	1 - 16	A	1 - 16	A
3	17 - 32	-	17 - 33	-
4	33 - 106	B	34 - 100	B
5	107 - 163	N	101 - 155	N
6	164 - 205	C-D	156 - 199	C-D
7	206 - 271	E	200 - 269	E

these domains particularly in light of the current evolutionary theories based on a role for introns in maintaining such integrity [45]. Alternatively, the conservation may simply result from the remnants of a common ancestor for the two genes.

The recent isolation and sequencing of a highly homologous (65% homology) murine counterpart to the human proIL-1β cDNA (personal communication, P. Gray, Genentech, Inc.) demonstrates the existence of two murine proIL-1 genes analogous to those found in the human. If the hypothesis that proIL-1β was derived from the alpha gene via retrotransposition is valid, then examination of the murine gene sequence, when it is isolated, should similarly reveal remnants of this event. It is interesting to speculate on the phylogenetic age of the duplication event responsible for generating two proIL-1 genes. Judging from the degree of difference between these sequences, and the fact that the murine genome carries two similar genes, the event must predate the divergence (about 75 million years ago) of the human and murine lineages [46]. Consequently, the proIL-1 gene may be at least 100 million years old. Reports suggesting that reptiles and fish are physiologically responsive to IL-1 protein [47, 48], and that starfish may produce IL-1-like molecules [49], supports a notion that the origin of this gene may extend back 450 to 700 million years ago, to a time before the elaboration of a true immune system.

Potential for alternate transcript processing. Introns 2 and 3 each contain a splice acceptor site which is highly conserved as a direct repeat within exon 2 immediately adjacent to the intron 2 splice donor site (labeled as direct repeat **b** in Figure 4). This conservation occurs at homologous sites within the proIL-1α gene, and at no other locations within these two sequences.

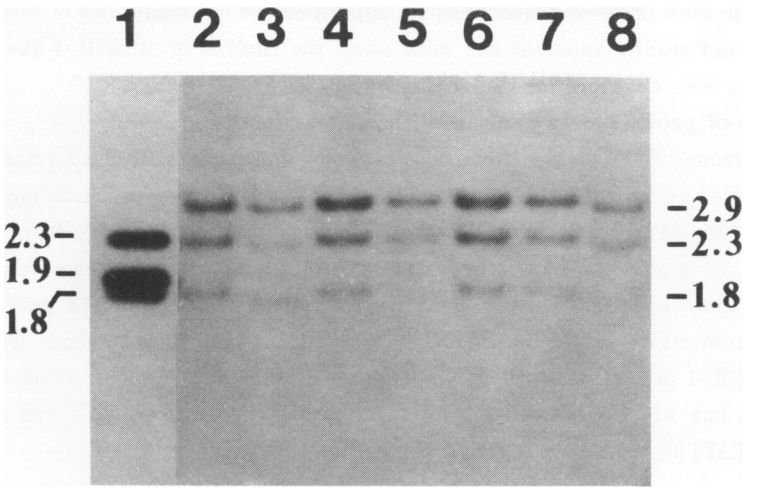


Figure 7. Southern blot analysis of *Hind* III restriction fragments of human genomic DNA derived from different cell sources which hybridize to radio-labeled proIL-1β cDNA. DNA samples in each lane are: (1) Colo 16, (2) leukocyte, (3) monocyte, (4) THP-1, (5) U937, (6) HeLa, (7) placenta, (8) bacteriophage clone BDC-454. The BDC-454 clone was digested with both *Hind* III and *Sal* I to remove the genomic insert from the λ vector sequence. The sizes of fragments are indicated in kilobase pairs.

```
TTTACACAG  Beta,  Intron 1 acceptor
||   ||||
TTATTACAG  Beta,  Exon 2
|||||||||
TTATTACAG  Beta,  Intron 2 acceptor
 | || |||
CTGTTGCAG  Alpha, Intron 1 acceptor
|||||  |||
CTGTTACAG  Alpha, Exon 2
 | |  ||||
TTTTGACAG  Alpha, Intron 2 acceptor

YYYYYNYAG  Mount acceptor consensus [41]
```

Such conservation suggests a role for these sequences, perhaps in alternate splicing. One plausible alternate splicing scheme, consistent with the sequence data, is a splice between exon 1 and the beginning of intron 2 with the splice acceptor sequence (TTATTACAG) at the end of exon 2. This would result in a new exon extending 65 nucleotides to a potential splice donor (GTGCTT at position 660) which could utilize the next available in-frame acceptor sequence at the beginning of exon 3. The resulting spliced message would code for a replacement of the first 16 amino acids of the proIL-1β

polypeptide for a different sequence of 13 amino acids. The significance of such a subtle change is not readily apparent and must await the analysis of other IL-1-like activities which have been described for various tissues [50, 51, 52, 53].

Presence of proIL-1 beta gene in different tissues

Human genomic DNA derived from different cell sources was analyzed for the presence of the proIL-1β gene using the proIL-1β cDNA as a hybridization probe. Figure 7 shows that all of the cell types examined have the same hybridization profile suggesting that there are no major gene rearrangements during differentiation of these tissues. In addition, there does not appear to be rearrangement of the proIL-1β gene in DNA isolated from either monocyte or THP-1 monocytic leukemia cells which are actively producing IL-1 in response to lipopolysaccharide stimulation. Another example is Colo 16, a cell line which constitutively produces the IL-1-like epidermal T cell activating factor (ETAF) [54] as well as a proIL1β-like mRNA [37].

ACKNOWLEDGEMENTS

We wish to thank the following people: Dr. M. Fenton for providing HeLa, U937, and THP-1 genomic DNA, Dr. K. Matsushima for THP-1 cells, Dr. E. Remold-O'Donnell for U937 cells, and Dr. D. Sauder for Colo 16 cells. Dr. W.A. Gilbert of the M.I.T. Whitaker College Computer Facility is acknowledged for assistance with computer software tools. We also wish to thank Drs. Sheldon Wolff, Alexander Rich, and Lee Gehrke for valuable contributions. This work was supported by a grant from Cistron Technology, Inc., Pine Brook, NJ and the Whitaker Health Sciences Fund, Cambridge, MA.

REFERENCES

1. Auron, P.E., Webb, A.C., Rosenwasser, L.J., Mucci, S.F., Rich, A., Wolff, S.M., and Dinarello, C.A. (1984) Proc.Natl.Acad.Sci.USA 81, 7907-7911.
2. March, C.J., Mosley, B., Larsen, A., Cerretti, P., Braedt, G., Price, V., Gillis, S., Henney, C.S., Kronheim, S.R., Grabstein, K., Conlon, P.J., Hopp, T.P., and Cosman, D. (1985) Nature 315, 641-647.
3. Lomedico, P.T., Gubler, U., Hellman, C.P., Dukovich, M., Giri, J.G., Pan, Y.E., Collier, K., Semionow, R., Chua, A.O., and Mizel, S.B. (1984) Nature 312, 458-462.
4. Furutani, Y., Notake, M., Yamayoshi, M., Yamagishi, J., Nomura, H., Ohue, M., Furuta, R., Fukui, T., Yamada, M., and Nakamura, S. (1985) Nucl. Acids Res. 13, 5869-5882.
5. Rosenwasser, L.J., Webb, A.C., Clark, B.D., Irie, S., Dinarello, C.A., Gehrke, L., Wolff, S.M., Rich, A., and Auron, P.E. (1986) Proc.Natl.Acad.Sci.USA 83, 5243-5246.
6. Dower, S.K., Call, S.M., Gillis, S., and Urdal, D.L. (1986) Proc.Natl.Acad.Sci. USA 83, 1060-1064.
7. Furutani, Y., Notake, M., Fukui, T., Ohue, M., Nomura, H., Yamada, M., and Nakamura, S. (1986) Nucl. Acids Res. 14, 3167-3179.

8. Rogers, J. (1983) Nature 306, 46.
9. Lee, M.G-S., Lewis, S.A., Wilde, C.D., and Cowan, N.J. (1983) Cell 33, 477-487.
10. Ueda, S., Nakai, S., Nishida, Y., Hisajima, H., and Honjo, T. (1983) EMBO J. 1, 1539-1544.
11. Battey, J., Max, E.E., McBride, W.O., Swan, D. and Leder, P. (1982) Proc.Natl.Acad.Sci.USA 79, 5956-5963.
12. Karin, M. and Richards, R.I (1982) Nature 299, 797-802.
13. Soares, M.B., Schon, E., Henderson, A., Karathanasis, S.K., Cate, R., Zeitlin, S., Chirwin, J., and Efstradiatis, A. (1985) Mol.Cell.Biol. 5, 2090-2103.
14. Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D., and Leder, P. (1982) Nature 296, 321-325.
15. Auron, P.E., Rosenwasser, L.J., Matsushima, K., Copeland, T., Dinarello, C.A., Oppenheim, J.J., and Webb, A.C. (1985) J.Mol.Cell.Immunol. 2, 169-177.
16. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T. (1978) Cell 15, 1157-1174.
17. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
18. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
19. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985) Plasmid 13, 31-40.
20. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc.Natl.Acad.Sci.USA 74, 5463-5467.
21. Staden, R. (1984) Nucl. Acids Res. 12, 505-519.
22. Staden, R. (1984) Nucl. Acids Res. 12, 521-538.
23. Staden, R. (1984) Nucl. Acids Res. 12, 551-567.
24. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucl. Acids Res. 12, 387-395.
25. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
26. Webb, A.C., Collins, K.L., Auron, P.E., Eddy, R.L., Nakai, H., Byers, M.G., Haley, L.L., Henry, W.M., and Shows, T.B. (1986) Lymphokine Res. 5, 77-85.
27. Webb, A.C., Auron, P.E., Rosenwasser, L.J., Mucci, S.F., Rich, A., Wolff, S.M., and Dinarello, C.A. (1985) Brit.J.Rheum. 24(suppl 1), 82-86.
28. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
29. Efstratiadis A, Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., and Proudfoot, N.J. (1980) Cell 21, 653-668.
30. Efstradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.R., Forget, B.G., Weissman, S.M., Slighton, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., and Proudfoot, N.J. (1980) Cell 21, 653-668.
31. Weiher, H., Konig, M., and Gruss, P. (1983) Science 219, 626-631.
32. Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984) Nature 308, 513-519.
33. Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986) Trends Biochem. Sci. 11, 20-23.
34. Gillies, S.D., Folsom, V., and Tonegawa, S. (1984) Nature 310, 594-597.
35. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
36. McLauchlan, J., Graffney, D., Whitton, J.L., and Clements, J.B. (1985) Nucl. Acids Res. 13, 1347-1368.

37. Auron, P.E. and Webb, A.C. (1987) Lymphokines , *in press*.
38. Sharp, P.A. (1983) Nature 301, 471-472.
39. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P. (1986) Ann. Rev. Biochem. 55, 1119-1150.
40. Jeffreys, A.J. and Harris, S. (1982) Nature 296, 9-10.
41. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
42. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc.Natl.Acad.Sci.USA 75, 4853-4857.
43. Keller, E.B. and Noon, W.A. (1984) Proc.Natl.Acad.Sci. USA 81, 7417-7420.
44. Reed, R. and Maniatis, T. (1985) Cell 41, 95-105.
45. Gilbert, W. (1978) Nature 271, 501.
46. Dayhoff,M.O. Ed., (1979) Atlas of Protein Sequence and Structure, Vol.5, Suppl.3,National Biomedical Research Foundation, Washington, D.C., .
47. Bernheim, H.A. and Kluger, M.J. (1977) J. Physiol. 267, 659-666.
48. Hamby, B.A., Huggins, E.M., Lachman, L.B., Dinarello, C.A., and Sigel, M.M. (1986) Lymphokine Res. 5, 157-161.
49. Donnely, J.J., Vogel, S.N., and Prendergast, R.A. (1985) Cellular Immunology 90, 408-415.
50. Luger, T.A., Stadler, B.M., Katz, S.I., and Oppenheim, J.J. (1981) J.Immunol. 127, 1493-1498.
51. Miossec, P., Cavender, D., and Ziff, M. (1986) J.Immunol. 136, 2486-2491.
52. Okada, M., Yoshimura, N., Ichimori, Y., Kishimoto, S., Nakai, S., Nishino, N., and Kishimoto, T. (1986) J.Immunol. 136, 1288-1294.
53. Fontana, A. and Grob, P.J. (1984) Springer Semin. Immunopathol. 7, 375-386.
54. Sauder, D.N., Mounessa, N.l., Katz, S.I., Dinarello, C.A., and Gallin, J.A. (1984) J. Immunol. 132, 828-832.