GENOMICS **29,** 490–495 (1995)

Structure and Evolution of the Human *IKBA* Gene

CARYN Y. ITO,* NILS ADEY,† VICTORIA L. BAUTCH,*,† AND ALBERT S. BALDWIN, JR.*,†,‡,[1]

**Curriculum in Genetics and Molecular Biology, †Department of Biology, and ‡Lineberger Comprehensive Cancer Research Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599*

Received February 17, 1995; accepted July 10, 1995

IκBα belongs to a gene family whose members are characterized by their 6–7 Ankyrin repeats, which allow them to interact with members of the Rel family of transcription factors. We have sequenced a human IκBα genomic clone to determine its gene structure. The human IκBα gene (*IKBA*) has six exons and five introns that span approximately 3.5 kb. This genomic organization is similar to that of other members of the Ankyrin gene family. The human *IKBA* gene shares similar intron/exon boundaries with the human *BCL3* and *NFKB2* genes, which is consistent with their conserved Ankyrin repeats. To examine further the evolutionary relationship between human IκBα and other members of its gene family, we performed a phylogenetic analysis. Although the resulting phylogenetic tree does not identify a common ancestor of the IκBα gene family, it indicates that this family diverges into two groups based on structure and function. © 1995 Academic Press, Inc.

INTRODUCTION

Interactions between transcription factors of the Rel family and members of the IκBα gene family have evolved to regulate genes mainly involved in immune and inflammatory responses (for review, see Grilli *et al.*, 1992; Nolan and Baltimore, 1992). The Rel family includes the oncoprotein Rel, RelA (p65), NF-κB1 (p50), NF-κB2 (p52), RelB, and the *Drosophila* morphogen, Dorsal. The Rel homology domain in the N-terminus of these transcription factors allows these proteins to bind DNA and to dimerize with certain members of this family. Various homo- and heterodimers of Rel family members preferentially interact with the Ankyrin repeats present in IκBα (Mad-3), NF-κB1 (p105), Bcl-3, and other proteins belonging to the IκBα family to mediate different functions in the cell (for review see Beg and Baldwin, 1993). IκBα interacts with p65 (Rel A), which together with p50 (NF-κB1) makes up NF-κB. This Rel–Ankyrin protein interaction causes the cytoplasmic retention of NF-κB, and therefore NF-κB is inhibited from binding to its target sequences and activating transcription in the nucleus (Baeuerle and Baltimore, 1988).

The precursor proteins, NF-κB1 (p105) and NF-κB2 (p100), contain a Rel homology domain in their N-termini and Ankyrin repeats in their C-termini. Thus, the Ankyrin repeats in the C-termini of these proteins are also capable of interacting with their Rel homology domains, forming an intramolecular Rel–Ankyrin protein complex that results in cytoplasmic compartmentalization (Rice *et al.*, 1992; Hatada *et al.*, 1992; Naumann *et al.*, 1993a,b; Mercurio *et al.*, 1993). Upon proteolytic processing, NF-κB1 and NF-κB2 lose their C-terminal Ankyrin repeats and are modified into p50 and p52, respectively. The processed proteins retain the N-terminal Rel homology domain and therefore the ability to dimerize with other Rel proteins and bind DNA, but unless they are coupled to IκBα, they are localized in the nucleus.

Bcl-3 has been identified only in mammalian species, and its function differs from those of other IκBα family members. Bcl-3 is found in the nucleus and preferentially interacts with NF-κB1 (p50) or NF-κB2 (p52) homodimers bound to DNA, and, unlike other Rel–Ankyrin interactions, this protein–DNA complex results in transcriptional activation (Wulczyn *et al.*, 1992; Bours *et al.*, 1993; Fujita *et al.*, 1993; Franzoso *et al.*, 1993).

In recent years, the genes encoding members of the IκBα family have been mapped by chromosomal translocations or by standard FISH mapping techniques (Ohno *et al.*, 1990; Neri *et al.*, 1991; LeBeau *et al.*, 1992; Ten *et al.*, 1992; Mathew *et al.*, 1993). The data show that the human genes in this family are not organized on a single chromosomal locus, but are scattered throughout the genome. This organization suggests that this gene family did not recently arise from simple duplication events. Furthermore, members of this family have been found in organisms as distantly related as flies, birds, and mammals, suggesting ancient origins. To understand better the evolution of this gene family, we have cloned and sequenced the human *IKBA* gene and compared it to other members of the Ankyrin

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. U08468.

[1] To whom correspondence should be addressed. Telephone: (919) 966-3652. Fax: (919) 966-3015.

0888-7543/95 $12.00

gene family. The *IKBA* gene contains six exons and five introns spanning 3.5 kb. Comparisons of the human *IKBA* gene with those encoding *BCL3* and *NFKB2* indicate that these genes share similar splice junctions and genomic organization. This gene organization is consistent with phylogenetic analysis that shows *IKBA* diverged from *BCL3* and *NFKB2* at least before mammalian and avian speciation, respectively.

MATERIALS AND METHODS

Isolation of the human IκBα genomic clone. A human placental genomic library was screened as described previously (Ito *et al.*, 1994). Briefly, 1×10^6 phage plaques were screened using a 5′ *Eco*RI/*Pst*I fragment from the human IκBα cDNA using the methods of Benton and Davis (1977). Two phage clones were isolated, and clone λ7-1 was further characterized by restriction enzyme mapping. The coding region of the gene was mapped to three *Sac*I fragments by Southern hybridization (Southern, 1975) using the IκBα cDNA clone as a probe. The three *Sac*I fragments were subcloned into pUC 19 for further analysis.

Mapping the intron/exon boundaries in the human IκBα gene. The following pairs of PCR primers were designed using sequences from the IκBα cDNA clone, 5′-AGACCTGGCCTTCCTCAAC-3′ and 5′-GTTGAGGAAGGCCAGGTCT-3′,5′-CCAACCAGCCAGAAATTC-3′ and 5′-CCAGCTCCCAGAAGTGCC-3′, 5′-CACTGCACACTGCCTAGCCC-3′ and 5′-GGGCTAGGCAGTGTGCAGTG-3′, 5′-CCCTCCCTGTAAATGGTGTAC-3′ and 5′-GTACACCATTTACAGGAGGG-3′. These primers were used to amplify regions of the cDNA and the genomic *Sac*I subclones. Sizes of the genomic and cDNA PCR products were compared on a 1% agarose gel run in 1× TAE to identify any introns. The introns were then sequenced in both directions using the primers mentioned above as well as the M13/pUC forward and reverse sequencing primers (NEB, Beverly, MA) using the Applied Biosystems automated sequencer at the UNC DNA sequencing facility.

Computer analysis of the IκBα genomic and amino acid sequences. The intron/exon boundaries in the human *IKBA* gene were compared to the intron/exon boundaries present in the human *BCL3* (McKeithan *et al.*, 1994) and *NFKB2* (p100/p52) (Fracchiolla *et al.*, 1993) genes using the PILEUP computer program (Feng and Doolittle, 1987). Human *NFKB1* (p105/p50) (Kieran *et al.*, 1990), mouse NF-κB1 (Ghosh *et al.*, 1990), chicken NF-κB1 (Capobianco *et al.*, 1992), human *NFKB2* (p100/p52) (Neri *et al.*, 1991), human *BCL3* (Ohno *et al.*, 1990), *Drosophila* Cactus (Geisler *et al.*, 1992), chicken IκBα (pp40) (Davis *et al.*, 1991), rat IκBα (Tewari *et al.*, 1992), human *IKBA* (Haskill *et al.*, 1991), and pig IκBα (de Martin *et al.*, 1993), *Drosophila* Notch (Wharton *et al.*, 1985), and human Tan-1 (Ellison *et al.*, 1991), obtained from the GenBank/EMBL database, were aligned using PILEUP. A phylogenetic tree based on amino acid similarity was generated using PAUP (phylogenetic analysis using parsimony, Swofford, 1990) computer programs. The analyzed sequences encompassed the Ankyrin repeats of the above members of the Ankyrin gene family. The resulting phylogenetic tree was replicated 100 times in a bootstrap confidence analysis.

RESULTS

Human IKBA Gene Structure

A lambda phage clone, λ7-1, containing 15 kb of recombinant insert DNA, was isolated from a human placental genomic library using a fragment of the IκBα cDNA as a probe. The coding region of the IκBα gene was localized to three *Sac*I fragments, which were then used to determine the structure of the IκBα gene by PCR analysis (shown in Fig. 1). The introns that were identified by PCR analysis were then sequenced to determine their splice junctions (Fig. 2). The human *IKBA* gene is relatively small, spanning 3.5 kb. It has six exons, ranging in size from 89 to 550 bp, and five introns ranging in size from 107 to 621 bp. The transcriptional start site was previously characterized downstream of a putative TATA box (Ito *et al.*, 1994), and the translation start site is located in the first exon. The translation stop site and polyadenylation signals are located in exon 6. The Ankyrin repeats are encoded by exons 2 through 5. Ankyrin repeats I and II are interrupted by introns 1 and 2, respectively, while repeats III through VI are encoded on single exons (see Fig. 3). Computer comparison of the genomic sequence of this gene to itself did not identify duplicated sequences (data not shown).

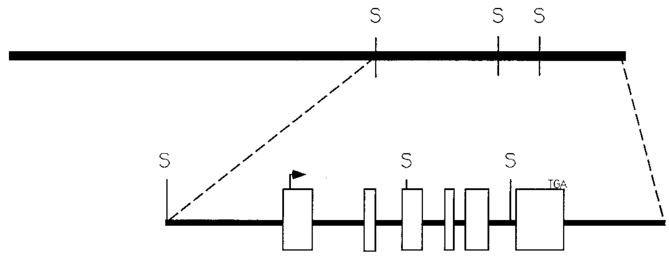

FIG. 1. A schematic diagram of the genomic IκBα clone, λ7-1. The enlarged diagram at the bottom represents the *IKBA* gene structure. The open boxes correspond to the exons, the arrow represents the translation start site, and the stop codon is shown in exon 6. S, *Sac*I restriction enzyme sites.

Because the first introns of several genes have been shown to contain important regulatory elements that serve as transcriptional enhancers, we analyzed the first intron of the *IKBA* gene for such sequences. Computer analysis of the first intron identified a putative NF-κB site (shown in Fig. 2). This NF-κB site is identical to one of the NF-κB sites in the HIV LTR enhancer (Nabel and Baltimore, 1987) and may serve to regulate this gene (see Discussion). Several putative regulatory elements were also identified; however, due to their weak homology to their respective consensus sequences, they are not considered significant.

Comparison of the IKBA Gene to Other Members of the Rel and IκBα Gene Families

To compare the gene organization of *IKBA* with other members of this gene family, we compared the nucleotide sequence of *IKBA* to the human genes *BCL3* and *NFKB2* using the PILEUP computer program. This analysis showed that several splice junctions are conserved among these genes, and exon sizes and the arrangement of their Ankyrin repeats were also conserved (Fig. 3). Exon/intron boundaries $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{3}{4}$ in *IKBA* correlate, respectively, to exon/intron boundaries $\frac{2}{3}$, $\frac{3}{4}$, and $\frac{4}{5}$ in *BCL3* and to boundaries $\frac{15}{16}$, $\frac{16}{17}$, and $\frac{17}{18}$ in *NFKB2*. An additional splice junction between ex-

```
gagctctgac cgaagtttgc tttattttgc ctgtaccaat tctaccaagt agatgaattc agtccatggc
ttgcaggctt ggaggtggag gtggagtgga gagcataaat tatatttctg caggtgaaac cagagcatag
tgagccgaaa cccctttcta cccaggtggc gtgaatgggt ccaactgcta ctgtccccca ggacaagaga
gcaaggaggg gcgggcagga tgggactacc ttgagcctcc cgcaggatgc ctggcaacta ctttcttagt
ggtccttaag gtccaatcgc gggttaaggc actaactaaa ctcattcttc agcctccttc tctaagttct
cctcgaattt gggtgcccag aagtaggctc acgatccttt ttctgcggga gcacaatgta ggtcagatag
cataaacgaa tagctactta tgaacacaat agctactctg ctattgcaag gtatccacaa ccaccactac
aaacattaac accttgcttt acttgggaaa caaaaaaaat catggtccta ttcagcagtt tccccataca
gggagcgttt gcccctcccc cagtcaacag ggctgttcat ccctaggaag tgatttgaga gttctccaag
gatttaggct ttcactcctc caaagctttc acaacttcta cctggcgggg gtgcgtgggg gggtgggggc
gaacgtaaaa gttcctttgc tgcaaagagc ctggtatagg cagaaacacc ggcgcggcct gcagcccccct
aaccacagtg cgtccttccc cttagaagtc tggggaaagc aaatccctac gcccagccat catttccact
cttgcgtttt caaaagatca aaaacggaaa ggaccggcag ttggcaaacc ccaaagaggg accgcccatc
aggtcggcgt ccttgggatc tcagcagccg acgaccccaa ttcaaatcga tcgtgggaaa ccccagggaa
agaaggctca cttgcagagg gacaggatta cagggtgcag gctgcaggga agtaccgggg ggagggggcc
tggtcggaag gactttccag ccactcggcg ctcatcaaaa agttccctgt ccgtgaccct agtggctcat
cgcagggagt ttctccgatg aaccccagct cagggtttag gcttcttttt ccccctagca gaggacgaag
ccagttctct ttttctggtc tgactggctt ggaaattccc cgagcctgac cccgccccag agaaatcccc
                +1
agccagcgtt tatagGGCGC CGCGGCGGCG CTGCAGAGCC CACAGCAGTC CGTGCCGCCG TCCCGCCCGC
CAGCGCCCCA GCGAGGAAGC AGCGCGCAGC CCGCGGCCCA GCGCACCCGC AGCAGCGCCC GCAGCTCGTC
CGCGCCATGT TCCAGGCGGC CGAGCGCCCC CAGGAGTGGG CCATGGAGGG CCCCCGCGAC GGGCTGAAGA
AGGAGCGGCT ACTGGACGAC CGCCACGACA GCGGCCTGGA CTCCATGAAA GACGAGGAGT ACGAGCAGAT
GGTCAAGGAG CTGCAGGAGA TCCGCCTCGA GCCGCAGGAG GTGCCGCGCG GCTCGGAGCC CTGGAAGCAG
CAGCTCACCG AGGACGGGGA CTCgtaagtg gcggaggccc ggggctgggg gtcgtcggga gggcgggacg
cgcgcggcct gggggtggga ggcgccgggc ctggagcccg agccggggga ccagcgatgc agccgggcgg
gcgccgcggc accccggggc tcctgcaccg ccccgaacgc tgcagggctg cagggccgcg cggggcaccg
gcccaactcg gagggggcggc ggctgggccc tggtggatgg cagacgtcgc cacacgcccc ctcccccccca
gacccgtcct ctaggccggg ggcggggcgg cagcggggac gcgaagtccc cggttgcata aggcggggcg
ggagtttctt gcccgctggc gggcgccgag cagccgggag gggaagtaca ggtcgttccg agctggcggg
agggcggggc ggtctggccg gcgcccgccg gggccgcggg gggagggggg cggaaagtcc ctgggcgcct
gccaggaaca ctcagctcat aataacctcg cggaaaacac cgcggcctcg gcctccagaa accccggcct
tgcgcaatcc cccgcagccc gcgcctccct gggccccacg cggtgcactc accaccccctg gggtttttcc
ctctcttccc cacagGTTCC TGCACTTGGC CATCATCCAT GAAGAAAAGG CACTGACCAT GGAAGTGATC
CGCCAGGTGA AGGGAGACCT GGCCTTCCTC AACTTCCAGA ACAACCTGCA GCAGgtgcgc cgcttgcctg
gcccgggttc tctctgaccc tgggacgtag ctgatgtagc agagtcaccc cagatccttt ctgaattcag
ggccactgag cactattcac cctcaccttt tacttcacat cagcccacat cctagagagt gaaggaaatt
ccactgattt ggtgaggtct ttatgaccca cctggagcct ctgctatttg ccagccctcc ccaccccccct
gtctaggagg agcagcaccc aaccaggaga cacgggttga ggggaactcg gggtgtgggt ttggtccatg
gcttactttc tctggtctct cttgcattcg tagACTCCAC TCCACTTGGC TGTGATCACC AACCAGCCAG
AAATTGCTGA GGCACTTCTG GGAGCTGGCT GTGATCCTGA GCTCCGAGAC TTTCGAGGAA ATACCCCCCT
ACACCTTGCC TGTGAGCAGG GCTGCCTGGC CAGCGTGGGA GTCCTGACTC AGTCCTGCAC CACCCCGCAC
CTCCACTCCA TCCTGAAGGC TACCAACTAC AATGgtatgt ctgcctccct gccctgcccc accccctcgg
agggcaggtg acgtggagaa ggggcaggtg ggccaactta aggagtccag gcaagagctt aaactcctaa
catttggaag gttgagaaaa tatgtgtgca aagtgcctat tggatgcctt tataaagttc tttcagaacc
cagactgtgg gttcttaaaa ttcagaagat aagactactt gttccataaa agaataggtg aaaggagtga
gggttgaaac aggtggttat actttttttcc ttttgttctt ccagGCCACA CGTGTCTACA CTTAGCCTCT
ATCCATGGCT ACCTGGGCAT CGTGGAGCTT TTGGTGTCCT TGGGTGCTGA TGTCAATGCT CAGgttggtg
cttcctgccc ccgacgcact gagtcaggct cctcgtgctt atgttgtgag cagaaattcc aaatgcagcc
ataagcatct caaattcctt ttggtgtcag GAGCCCTGTA ATGGCCGGAC TGCCCTTCAC CTCGCAGTGG
ACCTGCAAAA TCCTGACCTG GTGTCACTCC TGTTGAAGTG TGGGGCTGAT GTCAACAGAG TTACCTACCA
GGGCTATTCT CCCTACCAGC TCACCTGGGG CCGCCCAAGC ACCCGGATAC AGCAGCAGCT GGGCCAGCTG
ACACTAGAAA ACCTTCAGAT GCTGCCAGAG AGTGAGGATG AGGAGAGCTA TGACACAGAG TCAGAGTTCA
CGGAGTTCAC AGAGGACGAG gtgagtctgt gaactccttc ggcgctctaa ctaatgaggt gccattccct
tcaccctccc aggcccctag agctgctcct tatcagaggg gtatctacat aatgagtctc tcaaatttct
gtgcataacc agtatcccaa gaatgtacct gccccccccc ttttttttaat agcttactct tttttttaaag
agaagcatta aaggcaaact ccctcagcct gtaaagttca ttatatttgg gctatggaga atggagtcca
agagttattt ccagtagtgg cctccccatc ccggtagctt ggcagagctc cgctccggaa gcttaacgtg
tctttttttcc ccttgtttgc agCTGCCCTA TGATGACTGT GTGTTTGGAG GCCAGCGTCT GACGTTATGA
GTGCAAAGGG GCTGAAAGAA CATGGACTTG TATATTTGTA CAAAAAAAAA GTTTTATTTT TCTAAAAAAA
GAAAAAAGAA GAAAAAATTT AAAGGGTGTA CTTATATCCA CACTGCACAC TGCCTAGCCC AAAACGTCTT
ATTGTGGTAG GATCAGCCCT CATTTTGTTG CTTTTGTGAA CTTTTTGTAG GGGACGAGAA AGATCATTGA
AATTCTGAGA AAACTTCTTT TAAACCTCAC CTTTGTGGGG TTTTTGGAGA AGGTTATCAA AAATTTCATG
GAAGGACCAC ATTTTATATT TATTGTGCTT CGAGTGACTG ACCCCAGTGG TATCCTGTGA CATGTAACAG
CCAGGAGTGT TAAGCGTTCA GTGATGTGGG GTGAAAAGTT ACTACCTGTC AAGGTTTGTG TTACCCTCCT
GTAAATGGTG TACATAATGT ATTGTTGGTA ATTATTTTGG TACTTTTATG ATGTATATTT ATTAAAGAGA
TTTTTACAAA TG
```

FIG. 2. Genomic sequence of the human *IKBA* gene. The 5′ flanking sequences are in lowercase letters, the exons are denoted by capital letters, and the introns are in boldface, lowercase letters. The 5′ sequence containing three NF-κB sites and the transcription start site have been previously characterized (Le Bail *et al.*, 1993; Ito *et al.*, 1994). The three NF-κB sites in the promoter and the one site in the first intron are in boldface type and are underlined. The transcription start site is indicated by +1, and the translation start site and termination site are underlined.

ons 4 and 5 in *IKBA* is conserved between exons 5 and 6 in *BCL3*, and the splice site between exons 6 and 7 in *BCL3* is conserved between exons 18 and 19 in *NFKB2*. Since *BCL3* shares four splice junctions with *IKBA* and *NFKB2*, which share three splice junctions, the data suggest that these genes evolved from a common ancestor which, like *BCL3*, contained all five of these splice junctions.

To explore further the evolution of the IκBα gene family, a phylogenetic tree was constructed based on the Ankyrin repeats of protein sequences of a variety of members of the IκBα family. There were no significant differences between the phylogenetic trees using either the entire protein sequences or the regions encompassing the Ankyrin repeats (data not shown). The maximum parsimony method that relies on shared differences in amino acid sequences was chosen to construct the phylogenetic tree. To root the tree, the *Drosophila* Notch protein and its human homologue Tan-1 were used in this analysis. Notch also contains six Ankyrin repeats that were shown to be significantly similar to the human *BCL3* gene (Ohno *et al.*, 1990) and other members of the IκBα gene family (Nolan and Baltimore, 1992). In a distance matrix analysis comparing the percentage similarity of amino acids, Notch and Tan-1 are least similar to the members of the IκBα family (compare rows 1 and 2 with 3–14 in Table 1) and were therefore chosen as the outgroup to root the tree.

The phylogenetic tree (Fig. 4) shows that the Ankyrin repeats in Notch and Tan-1 diverged from the Ankyrin repeats in other family members before the speciation of vertebrates and invertebrates. This tree also shows that *IKBA*, *NFKB1*, and *NFKB2* diverged from each other before mammalian and avian radiation because there exists both a mammalian and a chicken homologue for these proteins. Because mouse Bcl-3 is the only homologue of human *BCL3* cloned at this time, the divergence of Bcl-3 from the others can be traced back only to the speciation of rodents.

This phylogenetic analysis groups the proteins into two major clades according to their structure and function. Cactus and the IκBα homologues are grouped together in one clade, and these two cytoplasmic inhibitors preferentially interact with the transcriptional activators, Dorsal and the RelA (p65) subunit of NF-κB, respectively. The IκBα proteins seem to be highly conserved, especially between the rat, pig, and human homologues, probably contributing to the low bootstrap values. The other major clade includes Bcl-3 together with NF-κB1 and 2. These proteins associate with NF-κB1 (p50) or NF-κB2 (p52) and seem to be grouped accordingly. This grouping suggests that the Ankyrin repeats present in NF-κB1 and 2 evolved to interact with their own Rel domain, p50 and p52, respectively. Bcl-3, which interacts with p50 as well as with p52, diverges from NF-κB1 and 2 and is placed in a separate group. However, the branch between Bcl-3 and NF-κB1 and 2 has a relatively low bootstrap value based on a 50% minimum consensus and therefore should not be considered absolute.

DISCUSSION

To better understand the evolution of the IκBα gene family, we have characterized the *IKBA* genomic organization and compared its gene structure to other members of this gene family. When the genomic sequence of *IKBA* is compared to itself, there are no duplicated sequences in the human *IKBA* gene. Moreover, the sequences of its Ankyrin repeats are more similar to their respective counterparts in other family members than to other repeats within the gene. The second Ankyrin

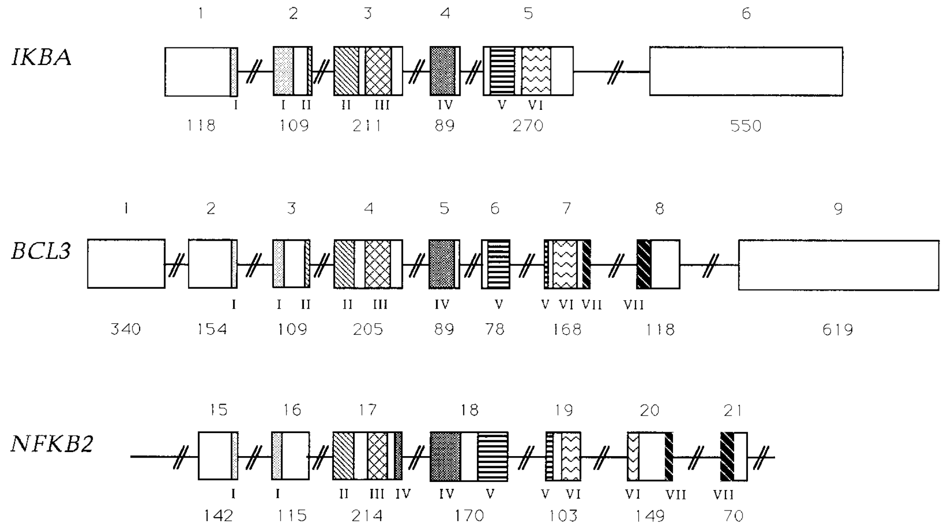

FIG. 3. Structure comparison of the human *IKBA*, *BCL3*, and *NFKB2* genes. Gene structures were compared using the PILEUP computer program. The boxes represent the exons numbered above them. The numbers below the exons correspond to their size in basepairs. The shaded boxes within the exons represent the Ankyrin repeats, designated by roman numerals.

repeat is the most conserved in this gene family, and the pattern of these repeats is also conserved (Nolan and Baltimore, 1992). The conservation of the Ankyrin repeats is consistent with our finding that the human *IKBA* gene shares four of the five intron/exon boundaries with the human *BCL3* gene, and three of these intron/exon boundaries are also observed in the gene encoding human NF-κB2. The conservation of the splice sites in this gene family suggests that these Ankyrin repeats may have arisen from duplications in a common ancestor prior to divergence of this gene family.

To determine how the Ankyrin gene family might have evolved, a phylogenetic tree was constructed. The resulting tree separates this family into clades according to structure and function. Therefore, while these proteins share a basic three-dimensional structure as predicted by their shared Ankyrin repeat pattern and sequence (Bork, 1993), a possible evolutionary scenario based on this phylogenetic tree could be that subtle differences in the amino acid substitutions in the Ankyrin repeats and flanking sequences occurred throughout evolution, which contributed to their specificity of interaction with various members of the Rel family. The Rel transcription factors diverge into two groups that include NF-κB1 and NF-κB2 in one clade and Dorsal, c-rel, and RelA (p65) on another (Schmid *et al.*, 1991). Likewise, the IκBα gene family also diverges into a group that interacts with either or both NF-κB1 and 2 and another group that interacts with

TABLE 1

Distance Matrix Analysis Comparing the Percentage Amino Acid Identities over the Length of the Proteins to Determine the Protein with the Least Percentage Sequence Similarity to Be Used as an Outgroup to Root the Phylogenetic Tree

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.0000	0.8201	0.3718	0.3621	0.3609	0.3722	0.3609	0.3722	0.3797	0.4000	0.3915	0.3872	0.3830	0.3410
2		1.0000	0.3419	0.3319	0.3640	0.3808	0.3682	0.3766	0.3975	0.3787	0.3574	0.3617	0.3532	0.3682
3			1.0000	0.9009	0.5299	0.5299	0.5043	0.5940	0.5214	0.4316	0.4231	0.4316	0.4103	0.4444
4				1.0000	0.5431	0.5388	0.5259	0.5560	0.5345	0.4483	0.4483	0.4483	0.4181	0.4569
5					1.0000	0.9560	0.8388	0.5421	0.6044	0.5191	0.5149	0.5106	0.4851	0.4291
6						1.0000	0.8498	0.5438	0.6022	0.5277	0.5191	0.5191	0.4979	0.4291
7							1.0000	0.5458	0.6007	0.5106	0.5021	0.5021	0.4979	0.4406
8								1.0000	0.6751	0.4766	0.4723	0.4809	0.4596	0.4674
9									1.0000	0.5234	0.5106	0.5021	0.4979	0.4598
10										1.0000	0.9617	0.9532	0.8553	0.5234
11											1.0000	0.9404	0.8468	0.5191
12												1,0000	0.8426	0.5064
13													1.0000	0.5021
14														1.000

Note. 1, *Drosophila* Notch; 2, human Tan-1; 3, human Bcl-3; 4, mouse Bcl-3; 5, mouse NF-κB1; 6, human NF-κB1; 7, chicken NF-κB1; 8, human NF-κB2; 9, chicken NF-κB2; 10, pig IκB-α; 11, human IκB-α; 12, rat IκB-α; 13, chicken IκB-α; 14, *Drosophila* Cactus.

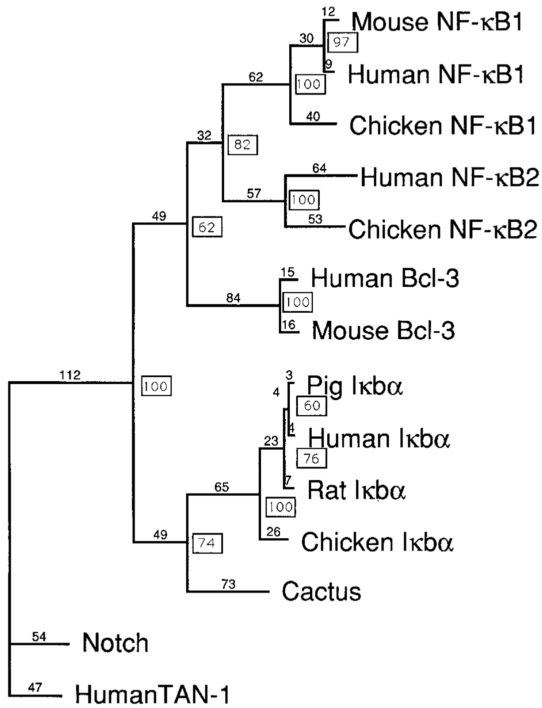

FIG. 4. Phylogenetic analysis of the IκBα gene family. A phylogenetic tree was generated by parsimony analysis of an alignment of the Ankyrin repeats. Notch and Tan-1 were defined as the outgroups based on the distance matrix analysis and were then used to root the tree (see Table 1). The branch lengths are proportional to the number of amino acid differences indicated by the numbers above the branches. The boxed values at each branch point were the results of a bootstrap analysis and represent the number of times that a particular node was obtained in 100 recalculations.

Dorsal, c-rel, and RelA. Thus, the divergence of the IκBα family is reflective of their interactions with the Rel family and resembles the divergence of the Rel family.

Other evolutionary clues to help identify a progenitor and understand the evolution of this gene family might be present in the 5′ and 3′ flanking regions or introns sequences. NF-κB sites have been identified in the porcine, murine, and human IκBα promoters and shown to regulate inducible gene expression (de Martin *et al.*, 1993; Chiao *et al.*, 1994; Le Bail *et al.*, 1993; Ito *et al.*, 1994). The first intron of the IκBα gene also contains a putative NF-κB site. This site is identical to the NF-κB site found in the HIV LTR, which has been previously characterized to confer UV-inducible transcription (Stein *et al.*, 1989). Therefore, this site may function as a transcriptional enhancer. The IκBα promoter is relatively weak (personal observations) and may require an enhancer to augment its activity. NF-κB and IκBα have been linked in an autoregulatory loop (reviewed in Beg and Baldwin, 1993), and this intronic NF-κB site may also participate in the regulation of the IκBα gene. The intron sequences of the IκBα gene in other species or other family members is not available for comparison. The finding of this NF-κB site in the introns of the IκBα gene in other species or other family members would indicate a functional role for this putative regulatory element. The intron sizes may also serve as clues to link family members and further elucidate the evolution of this gene family. The events involved in the evolution of the Ankyrin gene family and their roles in *Drosophila* axis formation and mammalian hematopoietic function and development may be clarified as more family members are discovered and their gene structures, functions, 5′ and 3′ flanking regions, and regulatory elements are elucidated.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Timothy McKeithan for sharing the Bcl-3 gene structure with us before publication. We also thank Dr. R. Kole, Dr. M. Edgell, and Dr. W. Stanford for their helpful discussions. We thank Dr. B. Kay for the use of his computers to run the sequence comparison programs. C.Y.I. was supported by a NIH training grant to the Curriculum in Genetics and Molecular Biology. This work was also supported by NIH Grants CA52515 and AI35098.

REFERENCES

Baeuerle, P. A., and Baltimore, D. (1988). IκB: A specific inhibitor of the NF-κB transcription factor. *Science* **242:** 540–546.

Beg, A. A., and Baldwin, A. S., Jr. (1993). The IκB proteins: Multifunctional regulators of Rel/NF-κB transcription factors. *Genes Dev.* **7:** 2064–2070.

Benton, W. D., and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* **196:** 180.

Bork, P. (1993). Hundreds of Ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally. *Proteins* **17:** 363–374.

Bours, V., Franzoso, G., Azarenko, V., Park, S., Kanno, T., Brown, K., and Siebenlist, U. (1993). The oncoprotein Bcl-3 directly transactivates through κB motifs via association with DNA-binding p50B homodimers. *Cell* **72:** 729–739.

Capobianco, A. J., Chang, D. D., Mosialos, G., and Gilmore, T. D. (1992). p105, the NF-κB p50 precursor protein, is one of the cellular proteins complexed with the v-Rel oncoprotein in transformed chicken spleen cells. *J. Virol.* **66:** 3758–3767.

Chiao, P. J., Miyamoto, S., and Verma, I. M. (1994). Autoregulation of IκBα activity. *Proc. Natl. Acad. Sci. USA* **91:** 28–32.

Davis, N., Ghosh, S., Simmons, D. L., Tempst, P., Liou, H-C. M., Baltimore, D., and Bose, H. R., Jr. (1991). Rel-associated pp40: An inhibitor of the Rel family of transcription factors. *Science* **253:** 1268–1271.

de Martin, R., Vanhove, B., Cheng, Q., Hofer, E., Csizmadia, V., Winkler, H., and Bach, F. (1993). Cytokine-inducible expression in endothelial cells of an IκBα-like gene is regulated by NF-κB. *EMBO J.* **12:** 2773–2779.

Ellison, L. W., Bird, J., Soreng, D. C., and Sklar, J. (1991). TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66:** 649–661.

Feng, D. F., and Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25:** 351–360.

Fracchiolla, N. S., Lombardi, L., Salina, M., Migliazza, A., Baldini, L., Berti, E., Cro, L., Polli, E., Maiolo, A. T., and Neri, A. (1993). Structural alterations of the NF-κB transcription factor lyt-10 in lymphoid malignancies. *Oncogene* **8:** 2839–2845.

Franzoso, G., Bours, V., Azarenko, V., Park, S., Tomita-Yamaguchi, M., Kanno, T., Brown, K., and Siebenlist, U. (1993). The oncoprotein Bcl-3 can facilitate NF-κB-mediated transactivation by removing inhibiting p50 homodimers from select kappa-B sites. *EMBO J.* **12:** 3893–3901.

Fujita, T., Nolan, G. P., Liou, H. C., Scott, M. L., and Baltimore, D. (1993). The candidate proto-oncogene *bcl-3* encodes a transcriptional coactivator that activates through NF-κB p50 homodimers. *Genes Dev.* **7:** 1354–1363.

Geisler, R., Bergmann, A., Hriomi, Y., and Nuesslein-Volhard, C. (1992). Cactus, a gene involved in dorsal ventral pattern formation of Dorsophila, is related to the IκB gene family of vertebrates. *Cell* **71:** 613–621.

Ghosh, S., Gifford, A. M., Riviere, L. R., Tempst, P., Nolan, G., and Baltimore, D. (1990). Cloning of the p50 DNA binding subunit of NF-κB: Homology to rel and dorsal. *Cell* **62:** 1019–1029.

Grilli, M., Chiu, J-S., and Lenardo, M. J. (1992). NF-κB and Rel-participants in a multiform transcriptional regulatory system. *Int. Rev. Cytol.* **143:** 1–62.

Haskill, S., Beg, A. A., Tompkins, S. M., Morris, J. S., Yurochko, A. D., Sampson-Johannes, A., Mondal, K., Ralph, P., and Baldwin, A. S., Jr. (1991). Characterization of an immediate-early gene induced in adherent monocytes which encodes IκB-like activity. *Cell* **65:** 1281–1289.

Hatada, E. N., Neiters, A., Wulczyn, F. G., Naumann, M., Meyer, R., Nucifora, G., McKeithan, T. W., and Scheidereit, C. (1992). The ankyrin repeat domains of the NF-κB precursor p105 and the protooncogene bcl-3 act as specific inhibitors of NF-κB DNA binding. *Proc. Natl. Acad. Sci. USA* **89:** 2489–2493.

Ito, C. Y., Kazantsev, A. G., and Baldwin, A. S., Jr. (1994). Three NF-κB sites in the IκBα promoter are required for induction of gene expression by TNFα. *Nucleic Acids Res.* **22:** 3787–3792.

Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J. S., Lottspeich, F., Le Bail, O., Urban, M. B., Kourilsky, P., Baeuerle, P. A., and Israel, A. (1990). The DNA binding subunit of NF-κB is identical to factor KBF-1 and homologous to the *rel* oncogene product. *Cell* **62:** 1007–1018.

Le Bail, O., Schmidt-Ullrich, R., and Israel, A. (1993). Promoter analysis of the gene encoding the IκBα/MAD3 inhibitor of NF-κB: Positive regulation by members of the rel/NF-κB Family. *EMBO J.* **12:** 5043–5049.

LeBeau, M. M., Ito, C. Y., Cogswell, P., Espinosa, R., Fernald, A. A., and Baldwin, A. S., Jr. (1992). Chromosomal localization of the Genes Encoding the p50/p105 subunits of NF-κB and IκB/MAD-3 inhibitor of the NF-κB to 4q24 and 14q13, respectively. *Genomics* **14:** 529–531.

Mathew, S., Murty, V. V. V. S., Dalla-Favera, R., and Chaganti, R. S. K. (1993). Chromosomal localization of genes encoding the transcription factors c-rel, NF-κB p50, NF-κB p65, and lyt-10 by fluorescence in situ hybridization. *Oncogene* **8:** 191–193.

McKeithan, T. W., Ohno, H., Dickstein, J., and Hume, E. (1994). Genomic structure of the candidate proto-oncogene *BCL3. Genomics* **24:** 120–125.

Mercurio, F., DiDonato, J. A., Rosette, C., and Karin, M. (1993). p105 and p98 precursor proteins play an active role in NF-κB mediated signal transduction. *Genes Dev.* **7:** 705–718.

Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326:** 711–713.

Naumann, M., Wulczyn, F. G., and Scheidereit, C. (1993a). The NF-κB precursor p105 and the proto-oncogene product Bcl-3 are IκB molecules and control nuclear translocation of NF-κB. *EMBO J.* **12:** 213–222.

Naumann, M., Nieters, A., Hatada, E. N., and Scheidereit, C. (1993b). NF-κB precursor p100 inhibits nuclear translocation and DNA binding of NF-κB/rel factors. *Oncogene* **8:** 2275–2281.

Neri, A., Chang, C-C., Lombardi, L., Salina, M., Corradini, P., Maiolo, A. T., Chaganti, R. S. K., and Dalla-Favera, R. (1991). B cell lymphoma-associated chromosomal translocation involves candidate oncogene *lyt-10*, homologous to NF-κB p50. *Cell* **67:** 1075–1087.

Nolan, G., and Baltimore, D. (1992). The inhibitory Ankyrin and activator Rel proteins. *Curr. Opin. Genet. Dev.* **2:** 211–224.

Ohno, H., Takimoto, G., and McKeithan, T. W. (1990). The candidate proto-oncogene *bcl-3* is related to genes implicated in cell lineage determination and cell cycle control. *Cell* **60:** 991–997.

Rice, N. R., MacKichan, M. L., and Israel, A. (1992). The precursor of NF-κB p50 Has IκB-like functions. *Cell* **71:** 243–253.

Schmid, R. M., Perkins, N. D., Duckett, C. S., Andrews, P. C., and Nabel, G. J. (1991). Cloning of an NF-κB subunit which stimulates HIV transcription in synergy with p65. *Nature* **352:** 733–736.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Stein, B., Rahmsdorf, J., Steffen, A., Litfin, M., and Herrlich, P. (1989). UV-induced DNA damage is an intermediate in the UV-induced expression of human immunodeficiency virus type 1, clooangenase, *c-fos,* and metallothionein. *Mol. Cell. Biol.* **9:** 5169–5181.

Swofford, D. (1990). Mac/PAUP: Phylogenetic analysis using parsimony. *In* "Illinois Natural History Survey," University of Illinois, Champain-Urbana.

Ten, R. M., Paya, C. V., Israel, N., Le Bail, O., Mattei, M. G., Virelizier, J. L., Kourilsky, P., and Israel, A. (1992). The characterization of the promoter of the gene encoding the p50 subunit of NF-κB indicates that it participates in its own regulation. *EMBO J.* **11:** 195–203.

Tewari, M., Mohn, K. L., Yue, F. E., and Taub, R. A. (1992). Sequence of rat RL/IF-1 encoding IκB-like activity and comparison with related proteins containing Notch-like repeats. *Nucleic Acids Res.* **20:** 607.

Wharton, K. A., Johansen, K. M., Xu, T., and Tsakonas-Artavanis, S. (1985). Nucleotide sequence from the neurogenic locus Notch implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell* **43:** 567–581.

Wulczyn, F. G., Naumann, M., and Scheidereit, C. (1992). Candidate proto-oncogene bcl-3 encodes a subunit-specific inhibitor of transcription factor NF-κB. *Nature* **358:** 597–599.