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

 $I\kappa B\alpha$ 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 I κ B α genomic clone to determine its gene structure. The human $I\kappa B\alpha$ 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 $I\kappa B\alpha$ 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 $I\kappa B\alpha$ 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 $I\kappa B\alpha$ 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- κ B1 (p50), NF- κ B2 (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 $I\kappa B\alpha$ (Mad-3), NF- $\kappa B1$ (p105), Bcl-3, and other proteins belonging to the I κ B α family to mediate different functions in the cell (for review see Beg and Baldwin, 1993). I κ B α interacts with p65 (Rel A), which together with p50 (NF- κ B1) makes up NF-

 κ B. This Rel–Ankyrin protein interaction causes the cytoplasmic retention of NF- κ B, and therefore NF- κ B is inhibited from binding to its target sequences and activating transcription in the nucleus (Baeuerle and Baltimore, 1988).

The precursor proteins, NF- κ B1 (p105) and NF- κ B2 (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- κ B1 and NF- κ B2 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 $I\kappa B\alpha$, they are localized in the nucleus.

Bcl-3 has been identified only in mammalian species, and its function differs from those of other $I_{\kappa}B\alpha$ family members. Bcl-3 is found in the nucleus and preferentially interacts with NF- κ B1 (p50) or NF- κ B2 (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 $I\kappa B\alpha$ 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.

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 $I\kappa B\alpha$ 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/*PstI* fragment from the human $I\kappa B\alpha$ cDNA using the methods of Benton and Davis (1977). Two phage clones were isolated, and clone λ 7-1 was further characterized by restriction enzyme mapping. The coding region of the gene was mapped to three *SacI* fragments by Southern hybridization (Southern, 1975) using the I κ B α cDNA clone as a probe. The three *SacI* fragments were subcloned into pUC 19 for further analysis.

Mapping the intron/exon boundaries in the human $I_{k}B\alpha$ gene. The following pairs of PCR primers were designed using sequences from the $I_{k}B\alpha$ cDNA clone, 5'-AGACCTGGCCTTCCTCAAC-3' and 5'-GTTGAGGAAGGCCAGGTCT-3',5'-CCAACCAGCCAGAAAT-TC-3' and 5'-CCAGCTCCCAGAAGTGCC-3', 5'-CACTGCACACTG-CCTAGCCC-3' and 5'-GGGCTAGGCAGTGTGCAGTG-3', 5'-CCC-TCCCTGTAAATGGTGTAC-3' and 5'-GTACACCATTTACAGGA-GGG-3'. These primers were used to amplify regions of the cDNA and the genomic *Sacl* 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 $I\kappa B\alpha$ 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 NFкВ1 (Ghosh et al., 1990), chicken NF-кВ1 (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 I κ B α (pp40) (Davis et al., 1991), rat I κ B α (Tewari et al., 1992), human IKBA (Haskill et al., 1991), and pig IkBa (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, λ 7-1, containing 15 kb of recombinant insert DNA, was isolated from a human placental genomic library using a fragment of the I κ B α cDNA as a probe. The coding region of the I κ B α gene was localized to three *Sac*I fragments, which were then used to determine the structure of the I κ B α gene by



FIG. 1. A schematic diagram of the genomic $I \ltimes B \alpha$ clone, λ 7-1. 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, *SacI* restriction enzyme sites.

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).

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- κ B site (shown in Fig. 2). This NF- κ B site is identical to one of the NF- κ B 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 $I \ltimes B \alpha$ 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-

	cyaayuuuyu	cccaccccyc	cegeaceaac	cocucouuge	ugueguueee	ageocacygo
ttgcaggctt	ggaggtggag	gtggagtgga	gagcataaat	tatatttctg	caggtgaaac	cagagcatag
tgagccgaaa	cccctttcta	cccaggtggc	gtgaatgggt	ccaactgcta	ctgtccccca	ggacaagaga
qcaaqqaqqq	gcgggcagga	tgggactacc	ttgageetee	cgcaggatgc	ctggcaacta	ctttcttagt
atcettaag	gtccaatcgc	gggttaaggc	actaactaaa	ctcattcttc	agectectte	tctaagttct
ctccaattt	agatacceaa	aagtaggete	acqatectt	ttetgeggga	gcacaatgta	ggtcagatag
cataaaccaa	tagetactta	tgaacacaat	agetacteto	ctattocaao	gtatecacaa	ccaccactac
	cagecuceta	osttoreses	agocuococg	estagtosts	ttaggaggtt	toocostsos
aacattaac	accttgettt	acttgggaaa	caaaaaaaa	calggiccia	Lucageague	Leeccataca
gggagcgttt	gecectecce	cagtcaacag	ggetgtteat	cectaggaag	tgatttgaga	gttetecaag
gatttaggct	ttcactcctc	caaagctttc	acaacttcta	cctggcgggg	gtgcgtgggg	gggtgggggc
gaacgtaaaa	gttcctttgc	tgcaaagagc	ctggtatagg	cagaaacacc	ggcgcggcct	gcagccccct
aaccacagtg	cgtccttccc	cttagaagtc	tggggaaagc	aaatccctac	gcccagccat	catttccact
cttgcgtttt	caaaaqatca	aaaacqgaaa	ggaccggcag	ttggcaaacc	ccaaagaggg	accgcccatc
aggtcggcgt	cettoggate	tcagcagccg	acqaccccaa	ttcaaatcga	tcqtgggaaa	ccccaqqqaa
	ottagagaga	ancegatte	cagatacaa	actacaaaaa	antacconon	agaagagaga
lgaaggoooa	anatttaana	gacaggacca	etestesses	acttocotot	agetagogg	agtogotcat
Lygreggaag	gactuccag	ccaccoggeg	CLCalCadaa	ageceeerge	ccycyaccec	ageggeeeac
cgcagggagu	Licicegalg	aaccccagcc	cayyyttay	geneerten	ccccccayca	yayyacyaay
ccagttetet	ttttctggtc	tgactggctt	ggaaattccc	_egagectgac	cccgccccag	agaaateeee
	+1					
agccagcgtt	tatag G GCGC	CGCGGCGGCG	CTGCAGAGCC	CACAGCAGTC	CGTGCCGCCG	TCCCGCCCGC
CAGCGCCCCA	GCGAGGAAGC	AGCGCGCAGC	CCGCGGCCCA	GCGCACCCGC	AGCAGCGCCC	GCAGCTCGTC
CGCGCCATCT	TCCAGGCGGC	CGAGCGCCCC	CAGGAGTGGG	CCATGGAGGG	CCCCCGCGAC	GGGCTGAAGA
TOODOAGOA	ACTGGACGAC	CGCCACGACA	GCGGCCTGGA	CTCCATGAAA	GACGAGGAGT	ACGAGCAGAT
CTCAACCAC	CTGCAGGAGA	TCCGCCTCGA	GCCGCAGGAG	GTGCCGCGCG	GCTCGGAGCC	CTGGAAGCAG
CACCTCACCC	ACCACCCCCA	CTCatarata	accasacco	anagat agaa	atcatcaaa	00000000000000000
CAGCICACCO	AGGACGGGGA	cregeaageg	gtggaggttt	99990099999	geegeeggga	aggeggggggg
egegeggeet	ggggggggga	ggegeeggge	ecggageeeg	ageeggggga	ccagegacge	ageegggegg
gegeegegge	accccggggc	reargeaceg	coccegaacyc	cgeagggeeg	cagggeegeg	cygygcaccy
geceaacteg	gagggggggg	ggetgggeee	tggtggatgg	cagacgroge	cacacgeeee	CCCCCCCCC
gaccogtect	ctaggccggg	aacaaaacaa	cageggggae	gcgaagtccc	cggttgcata	aggegggggg
ggagtttctt	gcccgctggc	gggcgccgag	cageegggag	gggaagtaca	ggtcgttccg	agetggeggg
agggegggge	ggtctggccg	gegeeegeeg	gggccgcggg	gggagggggg	cggaaagtcc	<u>c</u> tgggcgcct
gecaggaaca	ctcagctcat	aataacctcg	cggaaaacac	egeggeeteg	gcctccagaa	accccggcct
tococaatee	cccccacccc	acacctecct	gggccccacg	cootgcacte	accaccecto	gggtttttcc
etetetece	CACAGGTTCC	TGCACTTGGC	CATCATCCAT	GAAGAAAAGG	CACTGACCAT	GGAAGTGATC
etetetteee	cacagGTTCC AGGGAGACCT	TGCACTTGGC	CATCATCCAT	GAAGAAAAGG	CACTGACCAT	GGAAGTGATC
ctctcttccc CGCCAGGTGA	cacagGTTCC AGGGAGACCT	TGCACTTGGC GGCCTTCCTC	CATCATCCAT AACTTCCAGA	GAAGAAAAGG ACAACCTGCA	CACTGACCAT GCAGgtgcgc	GGAAGTGATC cgcttgcctg
ctctcttccc CGCCAGGTGA gcccgggttc	cacagGTTCC AGGGAGACCT tctctgaccc	TGCACTTGGC GGCCTTCCTC tgggacgtag	CATCATCCAT AACTTCCAGA ctgatgtagc	GAAGAAAAGG ACAACCTGCA agagtcaccc	CACTGACCAT GCAGgtgcgc cagatccttt	GGAAGTGATC cgcttgcctg ctgaattcag
ctctcttccc CGCCAGGTGA gcccgggttc ggccactgag	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac	TGCACTTGGC GGCCTTCCTC tgggacgtag cctcaccttt	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat	CACTGACCAT GCAGgtgcgc cagatccttt cctagagagt	GGAAGTGATC cgcttgcctg ctgaattcag gaaggaaatt
ctctcttccc CGCCAGGTGA gcccgggttc ggccactgag ccactgatt	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcaccttt ttatgaccca	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat cctggagcct	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctattg	CACTGACCAT GCAGgtgege cagateettt ectagagagt ecageeetee	GGAAGTGATC cgettgectg ctgaattcag gaaggaaatt ccaccccct
ctctcttccc CGCCAGGTGA gcccgggttc ggccactgag ccactgattt gtctaggagg	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct agcagcaccc	TGCACTTGGC GGCCTTCCTC tgggacgtag cctcacctt ttatgaccca aaccaggaga	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat cctggagcct cacgggttga	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctattg ggggaactcg	CACTGACCAT GCAGgtgcgc cagateettt ectagagagt ccagecetee gggtgtgggt	GGAAGTGATC cgcttgcctg ctgaattcag gaaggaaatt ccaccccct ttggtccatg
ctctcttccc CGCCAGGTGA geccgggttc ggccactgag ccactgattt gtctaggagg gcttactttc	cacagGTTCC AGGGAGACCT tetetgaccc cactattcac ggtgaggtot agcagcaccc tetggtotot	TGCACTTGGC GGCCTTCCTC tgggacgtag cctcaccttt ttatgaccca aaccaggaga cttgcattcg	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat cctggagcot cacgggttga tagACTCCAC	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctatttg ggggaactcg TCCACTTGGC	CACTGACCAT GCAGgtgcgc cagatocttt octagagagt ccagocotoc gggtgtgggt TGTGATCACC	GGAAGTGATC cgcttgcctg ctgaattcag gaaggaaatt ccaccccct ttggtccatg AACCAGCCAG
ctetetteee CGCCAGGTGA geeegggtte ggeeaetgag eeaetgattt gtetaggagg gettaettte AAATTGCTGA	cacagGTTCC AGGGAGACCT tetetgaccc cactattcac ggtgaggtet agcagcaccc tetggtetet GGCACTTCTG	TGCACTTGGC GGCCTTCCTC tgggacgtag cctcaccttt ttatgaccca aaccaggaga cttgcattcg GGAGCTGGCT	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat cctggagcct cacgggttga tagACTCCAC GTGATCCTGA	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctatttg ggggaactcg TCCACTTGGC GCTCCGAGAC	CACTGACCAT GCAGgtgcgc cagatocttt octagagagt ccagocotoc gggtgtgggt TGTGATCACC TTTCGAGGAA	GGAAGTGATC cgottgcctg ctgaattcag gaaggaaatt ccaccccct ttggtccatg AACCAGCCAG ATACCCCCCT
ctetetteee CGCCAGGTGA geeegggtte ggeeetgatte gtetaggagg gettaettte AAATTGCTGA ACACCTTGCC	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct agcagcaccc tctggtctct GGCACTTCTG TGTGAGCAGG	TGCACTTGGC GGCCTTCCTC tgggacgtag cctcaccttt ttatgaccca aaccaggaga cttgoattcg GGAGCTGGCT GCTGCCTGGC	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat cacgggttga tagACTCCAC GTGATCCTGA CAGCGTGGGA	GAAGAAAAGG ACAACCTGCA agagtcaccc cagccacat ctgctattg ggggaactcg TCCACTTGGC GCTCCGAGAC GTCCTGACTC	CACTGACCAT GCAGgtgcgc cagatcettt cctagagagt ccagocetec gggtgtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC	GGAAGTGATC ogottgootg ctgaattcag gaaggaaatt ccacccccct ttggtccatg AACCAGCCAG ATACCCCCCCT CACCCCGCAC
ctototteec CGCCAGGTGA gecegggtte ggccactgatt gtctaggagg gcttaottte AAATTGCTGA ACACCTTGGC CTCCACTCCA	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct agcagcaccc tctggtctct GGCACTTCTG TGTGAGCAGG TCCTGAAGGC	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcacctt ttatgaccca aaccaggaga cttgcattcg GGAGCTGGCT GCTGCCTGGC TACCAACTAC	CATCATCCAT AACTTCCAGA ctgatgtagc tacttcacat catggagcet cacgggttga tagACTCCAC GTGATCCTGA CAGCGTGGGA AATCGtatgt	GAAGAAAAGG ACAACCTGCA agagtcaccc cagccacat ctgctatttg ggggaactcg TCCACTTGGC GCTCCGAGAC GTCCTGACTC ctgcctccct	CACTGACCAT GCAGgtgcgc cagatcettt ccagagagt ccagecetce gggtgtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC ggcctgccc	GGAAGTGATC cgcttgcctg ctgaattcag gaaggaaatt ccacccccct ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC accccctcgg
ctotottoco CGCCAGGTGA gcocctgagt ccactgattt gtctaggagg gcttaottto AAATTGCTGA ACACCTTGCC CTCCACTCCA	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct agcagcaccc tctggtotct GGCACTTCTG TGTGAGCAGG TCCTGAAGGC acgtqgagaa	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcacettt ttatgaccca accaggaga ottgoattog GGAGCTGGCT GCTGCCTGGC TACCAACTAC gqqqcaqqtq	CATCATCCAT AACTTCCAGA ctgatgtage tacttcacat cacgggttga tagACTCCAC GTGATCCTGA CAGCGTGGGA AATGgtatgt gqccaactta	GAAGAAAAGG ACAACCTGCA agagtcacoc cagcccacat ctgctatttg ggggaactcg TCCACTTGGC GTCCGAGAC GTCCTGACTC ctgcctcct aggagtccag	CACTGACCAT GCAGgtgcgc cagatocttt octagagagt ccageoctoc gggtgtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gccctgcccc gcaagagctt	GGAAGTGATC cgcttgcdg gaggaaatt ccaccccct ttggtccatg AACCAGCCAG ATACCCCCCT CACCCGCAC accccctgg aaactcctaa
ctotottece CGCCAGGTGA geoeggtte ggcaetgagt gtatacttg gttaettte AAATTGCTGA ACACCTTGCC CTCCACTCCA agggeaggt	cacagGTTCC AGGGAGACCT tctctgaccc cactattcac ggtgaggtct agcagcaccc tctggtctct GGCACTTCTG TGTGAGCAGG TCCTGAAGGC acgtggagaa attgagaaaa	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcaccttt ttatgaccca accaggaga cttgcattcg GGAGCTGGCT GCTGCCTGGC TACCAACTAC ggggcaggtg tatgtgca	CATCATCCAT AACTTCCAGA etgatgtage tacttcacat cacggggtega tagACTCCAC GTGATCCTGA CAGCGTGGGA AATGgtatgt ggccaactta aactgcctat	GAAGAAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctatttg ggggaactcg TCCACTTGGC GCTCCGAGACC GTCCTGACTC ctgcctccct aggagtccag tcgatgcctt	CACTGACCAT GCAGgtgcgc cagatottt cctagagagt ccagocotco gggtgtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gcactgccco gcaagagett tataaaqttc	GGAAGTGATC cgcttgcctg gaaggaatt ccaccccct ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC acccctcgg aaactoctag tttcagaacc
ctctttccc CGCCAGGTGA gcccgggtc gccactgag ccactgatt gtctaggagg gcttacttc AAATTGCTGA ACACCTTGCC CTCCACTCCA agggcaggtg catttggag catcttggag	cacagGTTCC AGGGAGACCT tetetgaccc cactatteac ggtgaggtot tetggtaccc tetggtatet GGCACTTCTG TGTGAGCAGG TCCTGAAGGC acgtggagaa gttggagaaa gttggagaaa	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcacettt ttatgaceca aaccaggaga cttgcattcg GGAGCTGGC TACCAACTAC ggggcaggtg ttagtgtgca	CATCATCCAT AACTTCCAGA etgatgtage tacttcacat cacgggttga tagACTCCAC GTGATCCTGA CAGCGTGGGA AATCGtatgt ggccaactta aagatgcctat	GAAGAAAAGG ACAACCTGCA agagtcaccc ctgctatttg ggggaactcg TCCACTTGGC GTCCTGACTC GTCCTGACTC ctgcctccct aggagtccag tggatgcctaaaa	CACTGACCAT GCAGgtgege cagatocttt octagagagt cagoctoo gggtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gccatgecco gcaagagett tataagtto acaatagotg	GGAAGTGATC cgcttgcctg ctgaattcag gaaggaaatt ccaccccct ttggtccatg AACCAGCCAG ATACCCCCCAG accccctcgg aaactcctaa tttcagaacc acagagtga
ctcttccc CGCCAGGTGA gcccgggttc ggccaggtgtc ggctactgatt ggttaggag gcttacttc AAATTGCTGA ACACCTTGCC CTCCACCTGCC CTCCACTCCA	cacagGTTCC AGGAGACCT tototgaccc cactattcac ggtgaggtot agcagcaccc GGCACTTCTG TGTGAGCAGG TCCTGAAGGC acqtggagaa gttgagaaaa gttgtagaaaa	TGCACTTGGC GGCCTTCCTC tgggacgtag cctacacttt ttatgaccoa aaccaggaga dtggattcg GGACTGGCT GCTGCCTGGC TACCAACTAC ggggcaggtg tatgtgtgca ttaggagat	CATCATCCAT AACTITCCAGA ctgatgtagc tacttcacat catggagctt tagACTCCAC GTGATCCTGA CAGCGTGGA AATGgtatgt aggccaactta aagtgoctat aagtcactt	GAAGAAAAGG ACAACTGCA agagtcaccc cagcccacat ctgctatttg ggggaaccg TCCACTTGGC GCTCCGAGAC GTCCTGACTC ctgcctccat aggatccag tggatgcctt gttccataaa ccagCCCACA	CACTGACCAT GCAGgtgcgc cagatcott cctagagat ccagocotcc ggdtgtggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gccatgccoc gcaagagctt tataaagttc agaatagttg cGTGTCTACA	GGAAGTGATC ogottgoctg ctgaattcag gaaggaaatt coaccccct ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC accccctag attccagaacc aaggagtga CTTAGCCTCT
ctctattece csccAsGTGA ggccaggtta ggccatgatg ggttaggag ggttacttto AAATTGTGA ACACCTTSCC CTCCACTCCA aggccaggtg catttggag catttggag gggttgaac	cacagGTTCC AGGAGACCT tottgaccc cactattcac ggtgagtct tagcagcaccc totggtotot GCCACTTCTG GCCACTTCTG TGTCACCACG TCCTGAAGGC acgtggagaa gttggagaaa gttggagaaa agttgggtat	TGCACTTGCC GGCCTTCCTC tgggacgtag octoacottt ttatgaccca aaccaggag cttgoattog GGACCTGCC TACCAACTAC gggcaggtg ttoggaggtg ttoggaggta cttcacactt	CATCATCCAT AACTITCCAGA ctgatgtag tacttcacat cacgggdcgt tagACTCCAC GTGATCCTGA CACCTGGGA AATCgtatgt ggccaacta aagtgoctat ttttgttctt ttttgttctt	GAAGAAAGG ACAACCTGCA agagtcaccc cagcccacat ctgctattg ggggaactcg TCCACTTGGC GCTCCGACAC CTCCGACAC ctgcctcaCAC ctgcctcaCAC ggagtcag tggatgcctt gttccataaa ccagCCACA	CACTGACCAT GCAGgtogo cagatocttt octagagagt ccagoctce ggtgtgggt TGTGATCACC TTTCCGAGAA AGTCCTGCAC gccatgccoc gcaagagctt tataaagtc agaataggtg CGTGTCTACA	GGAAGTGATC ogottgoctg ctgaattcag gaaggaaatt ccaccocct ttogtcoctg AACCAGCCAG ATACCAGCCAG ACACCCCCCT acoccotcgg aactoctaa tttcogaactoctaa tttcogagagtga CTTAGCCTCT
ctcttccc cccCAGGTGA gcccgggtc ggccactgag gcttactto AAATTCCTGA ACACCTTGCC CTCCAATCCA agggcaggtg catttggag caggctgaaac	cacagGTTCC AGGGAGACCT tettgacc catattcac ggtgaggtct agcagcacc tetggtetet GGCATTCTG TGTGAGCAGG TCTTGAGGCA gttgagaaa gttgagaaa gttgagaaa gttgagaaa gttgagaaa	TGCACTTGGC GGCCTTCCTC tgggacqtag octoacott ttatgaccoa aaccaggag ottgoatcg GGAGCTGGCT GCTGCCTGGC TACCACTAC tagggacagtg tatgtgtgca ttaggagat catttttcc CGTGGAGCTT	CATCATCCAT AACTTCCAGA Ctgatgtagc tacttcacat octggagott tagAATCCAC GTGATCCTGA CAGCGTGGGA AATGgtatgt aggccaactta aagtgoctat aagtgoctat ttttgttctt TTGGTGTCCT	CAAGAAAGG ACAACCTGCA agagtcacce cagcccacat ctgctattg ggggaactog TCCACTTGGC GTCCTGACTC GTCCTGACTC ctgcctccct gggatgcctt gttccataaa ccagGCCCACA	CACTGACCAT GCAGgtogto cagatocttt octagaogt ggtgtggg TGTGATCACC gcatgcoct gcasgagett tataagttc agaatagttg GGTGCTACA	GGAAGTGATC ogottgoctg otgaattcag gaaggaaatt ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC acococtogg aaactoctaa tttcagaacc aaggagtg CTTRACCTCT CACGttggtg
attattee cGCCAGGTGA ggeceggtte ggecactgatt ggttaggag gattactte AAATCATGCG ACACCTTGCC CTCCACTCCA aggecagatg cagactgtg gggttgaaa ATCATGGCT atteggec	cacagGTTCC AGGAGACCT tcttgacc catattcac ggtgagdct agcagcaccc tctggtctt GGCACTTCTG TGTGAGCAGG acgtggagaa gttgtagaaaa gttgttaaaa agttgttat ccrGGCAT ccgacgcat	TGCACTTGGC GGCCTTCCTC tgggacgtag octoacott ttatgaccoa accaggag ottgoattog GGAGCTGCT GCTGCCTGGC ggggcagtg ttoggagt ttoggagt actttttoc CGTGGAGCTT gggtaggc	CATCATCCAT ANCTTCCAGA ctgatgtagc tagtftagc tagtftagc tagACTCCAC GTGATCCTGA CACCCTGGA ANTGgtatgt ggccaactta aggctaatt ttttgttatt ttttgttatt ttttgttagt cacgtggt	CAAGAAAGG ACAACCTGCA agagtcacoc cagcocacat ctgotatttg ggggaactog TCCACTTGGC GCTCCGAGAC GTCCTGACTC ctgoctcoct aggagtccag tggagtoctt gttccataa ccagCCACA atgftgtgag TTGGGTGCCTGA	CACTGACCAT GCAGgtogg cagatoctt ccagacatt ggtqtgggt TGTGATCACC gcctgcacg gcctgcacg gcagagett tataagtt cGTGTCTACA gGCTGTCTCACA	GGAAGTGATC ogottgaattcag gaaggaaatt ttgqtccatg AACCAGCCAG AACCAGCCAG ATACCCCCCCCCCC CACCCCCCCCAC accocotcgg aaactcataa ttccagaacc acaggagtga CTTAGCTCCT ACGtggtgg aaafgagcoc
attattee GCCAGGTGA gcecgggta ggcaatgag gctaatta gtctaggag gcttactte AAATTGCTGA ACACCTTGCC CTCCACTCCA CTCCACTCCC CTCCACTCCA	cacagGTTCC AGGGAGACCT tetetgacce catattcac ggtgaggtet dgcagcacct tetggtetet GGCATTCTG TGTGAGCAGG TCTTGAAGGC TCTGAAGGC acgtggagaa gttettaaaa ggtgttat ACCTGGGCAT ccgaegcact ccgaegcact	TGCACTTGCC GGCCTTCCTC tgggacgtag octaaoctt ttatgaccaa accaggag ctggactag cdaCTGGCT GCTGCCTGC TACCAACTAC ggggcaggtg tatgtgca ttcaggagat tcaggagat ttaggtgcaggt	CATCATCCAT AACTTCCAGA ctgatgtag tacttcacat ootggagoot tagAATCCAC GTGATCCTGA CACCTGGGA AATGgtatgt aggccaactta aaggcotat aagactactt tttgttctt TTGGTGTCCT cotogtgott GAGCCTGTA	CAAGAAAGG ACAACCTGCA agagtcacce cagccacat ggggaactog TCCACTTGGC GTCCTGACTC ctgoctoct gtggtgcctt gtcotacac gtggtgcctt gtcotata gtccafGCTGA ATG6GTGGTCG ATG6CCGGCC	CACTGACCAT GCAGgtogo cagatodtt ocagocotto ggdygtggt TGTGATCACC gcotgcoot gcaagagott tataagtto GGTGTCTACA TGTCAATCCT GcGAATCCT GCCATCAC	GGAAGTGATC ogottgoatg ctgaattcag ccaceccect ttggtccatg AACCAGCCAG ATACCCCCA CACCCCGAC CACCCCGAC aacoccotcgg aactoctaa atttcagaacc aaggagtg aattgagcc CTTAGCCTCT CAGttggtg CTTCGCAGTGG CTTCGCAGTGG
ctcttecc cGCCAGGTGA gcccgGgtc ggccactgag gottacttc AAATTGCTGA ACACCTTGCC CTCCACTTGCC CTCCACTTGCC aggcgaggtg ggttgaac ATCCATGGCT dttactgocc ataggcatt	cacagGTTCC AGGGAGACCT AGGGAGACC catattcac ggtgggct agosgcaccc totggtott GGCACTTCTG TGTGAGCAGG gttggggaa gttgtagaaa gttgtgggtat ACCTGGGCAT ccgacgCact casattoctt TCCTGACCTG	TGCACTTGGC GGCCTTCCTC tgggacgtag cotoacott ttatgaccaa accaggaga cttgoattog GGACCTGGCT GCTGCCTGGC TGCCAACTAC gggcaggtg tatgtgtca ttcagaagat actttttcc CGTGGAGCTT ggtcaggct ttggtcagg ttggtcaggct	CATCATCCAT ANCTTCCAGA ctgatgtagc tagttcacat catgagget tagACTCCAC GTGATCCTGA CACCGTGGA AATGgtatgt ggccaatta aagtgcotat ttttgttatt tTTGGTGCTCCT catogtgott GAGCCCTGTA	CAAGAAAGG ACAACCTGCA agagtcacoc cagcccacat ttgcatttg ggggaactog TCCACTTGGC GCTCCGAGC GTCCTGACTC ctgcatcact agagtccag tggatgcctt gttccataa ccagCCACA TGGGTGCTGA ATGGCCGGAC TGGGGCTGAT	CACTGACCAT GCAGgtgcgc cagatottt cctgagagt TGTGATCACC TTTCGAGAA AGTCCTGCAC gcatgccc gcagagctt tataagtc agaatagtg CGTGTCTACA TGTCAATGCT Gagaattcc TGCCCTCACC	GGAAGTCATC cygattgatg gaggaatt ccasccccct ttgytcatg AACCAGCCAG ATACCCCCCCCC CACCCCCCCCCCCCC
attattae coccacquesta goccacquesta goccacquesta goccacquesta goctagaga gottaatta acacctreec crecactreec crecactrees aggotagatg gggttgaac Arccarageata Accaccacaa acttactgeec attactgeec	cacagGTTCC AGGGAGACCT tetetgaccc cactattcac ggtgaggtct agcaggtct GGCACTTCTG GGCACTTCTG TGTGAGCAGG TCCTGAAGGC agttggagaa gttgtgagaa ggtggtat ACCTGGGCAT ccgacgCact ccgacgCact ccCTACCAGC	TGCACTTGGC GGCCTTCCTC tgggacgtag cotoacott ttatgaccoa accaggaga cotogattog GGACCTGGC TACCAACTAC gggcaggtg tatgtgtgca tcagaagat tcagaagat ttggtgtag GTGTCACTCC TCACCTGGGG	CATCATCAAT ANCTTCCARA ctgatgtagc tacttcaat cacggagct tagACTCCAC GTGATCCTAC CACCTGGGA AATGgtatgt ggccaacta aagtgoctat aagtgoctat aagtgoctat cacggtgett CACCTGGTCCT Ctogtgett GAGCCCTGTA TGTTGAACTG	CAAGAAAGG AcAACCTGCA agagtcaccc cagcccaat TCCACTTGC CTCCGAGAC CTCCGAGAC CTCCGAGAC ctgoatocat aggagtcag tggatgocat gtgoatgocat gtgoatgocat aggagcocag ATGGCCGAC Atgttgtgag ATGGCCGGAC TGGGGCTGAT	CACTGACCAT GCACGtgege cagatoottt octgagaggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gccatgecce gcagagott tataagttc agataggtg GGTGTCTACA TGTCCATCAT TGCCCTTACA GTCAACGAC	GGAAGTGATC ogottgootg ctgaattcag ccacccccct ttggtccatg AACCAGCCAG ATACCGCCAG CACCCCGCAC acoccccgg aactoctaa attcagaac aaggagtg aatgoagcc CTGGCAGTGG TTACCTACCA GGGCCAGCTGG
cccccccccccccccccccccccccccccccccccccc	cacagGTTCC AGGGAGACCT tcttgacc catattcac ggtgaggtct agcaggcaccc totgttott GGCACTTCTG TCCTGAAGCAGG acttgagaa gttgsgaaa gttgsgaaa gttgsgaaa ACCTGGCAT ccaagcact TCCTGACCAGC CCTACCAGC	TGCACTTGGC GGCCTTCCTC tgggactag cctoacott ttatgacoa accaggaga ctggattog GGAGCTGGCT GCTGCCTGGC tatgtgtga ttoggaggtg ctatgtgtga tcagagat actttttco CGTGGAGCTT ggtcaggct ttggtgtoag GTGTCACTCC CCCCCGGGG GCTGCCACAGG	CATCCATCCAT ANCTTCCAGA ctgatgtagc tactgaggtagc tagACTCCAC GTGATCCTGA CAECCTGGA AATGgtatgt ggccaacta aagtgoctat aagtgoctat ttttgttctt tTTGGTGCTCCT cctogtgott GAGCCCTGTA TGTTGAAGTG CCGCCCAGGC	CAAGAAAGG ACAACTGCA agatcacc cagcccacat TCCACTTGCC GTCCGAGAC GTCCGAGAC GTCCTGACTC ctgcctccct agagtccag gtgatccag gtccataaa ccagCCACA atgttgtgag ATGGCCGGA ATGGCCGGA TGGGGCGGAT ACCCGGATAC	CACTGACCAT GCAGgtgcgc cagatottt octgagagt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gcatgccc gcagagctt tataagttc agaatagtg CGTGTCTCAC GGCATCACAGAG AGCAACAGAG	GGAAGTCATC cogottgootg ctgaattcag gaaggaatt ccacccccct ttggtccatg AACCAGCCAG ATACCCCCCC CACCCCGCAC cacccctcgg aactoctag attcagaacc aaggagtga CTTAGCCTC CAGgttggtg aatgcagcc CTCGCAGTGG TTACCTACCA GGGCCAGCTG TCACAGTTCA
ctcttecc cccCaGGTGA gcccgggttc ggccactgag ggctacttg gctacttc ANATTGCTGA ACACCTTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC AGgggttgaac ATCCATGCCT attaggactgag aggsttgaac attaggact attag attaggact attag attaggact attag att	cacagGTTCC AGGGAGACCT tetetgaeee cactatteae ggtgaggtet ageagcaece tetggtetet GCCACTTCTG TGTGAGCAGG TCCTGAAGGC acgtgggaa gttgagaaa gttgtaaaa gttgtaaaa gtgtgttat ACCTGGCATT ccgaegcaet ccaaateett TCCTGACTG CCCTACCAGC ACCTTCACAT	TGCACTTGGC GGCCTTCCTC tgggacgtag cotcacctt ttatgacca accaggaga GGACCTGGC TACCAACTAC gggcaggtg tatgtgca ttagtgtgca tcagagat ttggtgcagg GTGTCACTCC CCACCGGGG GCTGCCACAG gtgagtctgt	CATCATCAAT ANCTTCCARA ctgatgtagc tacttcaat cacggagct tagACTCCAC GTGATCCTAC CACCTGGGA ANTGgtatgt ggccaactta aagtgcotat aagtgcotat aagtgcotat aagtgcotat CACCTGGTCCT TGTGGTCCT CCCCCCAAGC AGTGAGGATG AGTGAGGATG AGTGAGGATG Gaactcattc	CAAGAAAGG AcAACCTGCA agagtcaccc cagcccacat TCCACTTGC CTCCTGACTC ctgoctocct aggagtcag ctgostgoctt gtgatgoctt gtgatgoctt gtgatgoctt gtGCCGCACA TGGGTGCTGA Atgttgtgag ATGGCCGGCC TGGGCCGACA AGGAGGTAA	CACTGACCAT GCACGtgege cagatoottt octgagaggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gcoctgecec gcagagett tataagttc agataggtg GGTGTCTACA TGTCAATGCT GGCAATGCT TGACAAGAG GTCAACAGAG GCCACACACG Gtaatgaggt	GGAAGTGATC ogottgoctg ctgaattcag cagoccocc ttggtcoatg AACCAGCCAG AACCAGCCAG AACCAGCCAG aactoctag attcagaace aaggagtg CTTAGCTGCTC CAGGtgGtG TTACCTACCAG CGGCCAGCTG TCACGAGTTCA GGGCAGCTG CAGAGTTCA
cccccacgagt gcccgggtta ggccactgag ggccactgag ggctactta ggttaggag gctactta arggcaggtg cagactgtg gggttgaaac antccArGeCT attagcagtc attagcatt ACCCTGCATAA GGGCTATTCT ACCATAGAT	cacagGTTCC AGGGAGACCT AGGGAGACCT ccttgacc catattcac ggtaggtct gGCACTTCTG TGTGAGCAGG gttggagaa gttggagtat ACCTGGCAT ccgagcact ccgagcact CCCTAACGG CCCTACCAGC ACCTTCACCTG CCCTACCAGC	TGCACTTGGC GGCCTTCCTC tgggactag cctoacott ttatgacca accagagag GGAGCTGGCT GCTGCCTGGC TACCAACTAC gggcaggtg tatgtgca ttatgtgca tctttttc CGTGGAGGCT gagtaggct ttggtptcag GTGTCACTGC CCTCCCAGGG gtgagtcgt	CATCATCCAT ANCTTCCAGA atgatgtagc tagttcacat catgaggtg tagACTCCAC GTGATCCTGA CAGCGTGGA AATGgtatgt ggccaacta aagtgcotat aagtgcotat ttttgttctt tTTGGTGGTCCT cctogtgctt TGGTGACTGA CAGCCCAAGC CGCCCAAGC CAGCACGATG gaactactt tatcagaggg	CAAGAAAGG ACAACTGCA aggtcacca cagcccaat TCCACTTGC GCTCCGAGAC GTCCGAGAC GTCCTGACTC GTCCGAGAC GTCCTGACTC TGGGGCCCA TGGGGCCCGA Atgttgtgag ATGGCCGAC ACCCGATAC ACCCGATAC AGGAGACGTA ggtactacat	CACTGACCAT GCAGgtggg cagatottt octgagaggt TGTGATCACC TTTCGAGGAA AGTOCTGCAC gcatgccco gcagaggtt tataaagtc agaatagtg CGTGTCTCAC GTGAATGCT CGCCTTCAC GTCAACAGG AGCAGCAGCT TGACACAGG ctaatgagtt	GGAAGTCATC cogottgoctg ctgattcag gaggaatt ccacccccct ttggtccatg AACCAGCCAG ATACCGCCAG ATACCGCCCT CACCCCGCAC CACCCCGCAC CTTAGCCCCT CAGgttggtg aatgcagcc CTCGCACTGG TTACCTACCA GGGCCACTGG TTACCACCA GGGCCACTGG TCACAGTTCA gcocttcct tcaatttct
ctcttecc GCCAGGTGA gcccgggtc ggccactgag gctacttc AAATTGCTGA ACACCTTGCC CTCCACTTGCC CTCCACTTGCC agggcaggtg gggttgaaac ATCCATGGCT attactgoca taggcattgag gggttgaaac ATCCATGGCAAAA GGGCTATTCT ACCACAAAA GGGCATTCAC tagcactca	cacagGTTCC AGGAGACCT AGGAGACC catattcac ggtgagdct agcaggacc totggtott GGCACTTCTG TGTGAGAGG aggtggaga gttgtagaaaa gttgttaaaa agttggtat ACCTGGCAT CCGGAGGCAT CCGGACCGC CCTACAGC ACCTCAGAT AGAGGACCAG aggtacccag ggtatcaca	TGCACTTGGC GGCGTTCCTC tgggacqtag cttaggacqtag ctagactag ctggactag ctggacagtag actgoattog GGAGCTGCT GCTGCCTGGC ggggcagtg ttoggagatg ttoggagatg ttoggagatg ttggtgcag GTGTCACTCCTCG GGCGCAGAG GCTGCCAGAG gtgggtctgt agtggtcag	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tactgaggtsga tagACTCCAC GTGATCCTGA CAGCGTGGA ANTGgtatgt ggccaacta aggccaacta ttttgttctt tttgttctt tttgttctt TGGTGGTCCT Catcagtgot GAGCCCTGTA AGTGAGGATG ggactccttc tatcagaggg gcccoccocc	CAAGAAAGG ACAACCTGCA agagtcacce cagcccacat ctgctatttg ggggaactcg TCCACTTGGC GTCCGACAC GTCCTGACTC ctgcatccat aggagtccag tggatgcctt gttccataa ccagCCACA atgttgtgag ATGGCCGGATC AGGAGCTGA ACCCGGATAC AGGAGGACTA Aggcgtctaa gtattacacat	CACTGACCAT GCAGgtogg cagatoctt ccagacatt ggtqtgggt TGTGATCACC gcctgccc gccagactt tataagtt cGTCTCTCCAC gccagagctt tataagtg CGTGTCTTCAC GGCATCTCTCC GGCAACAGAG GCCAACAGAG GCCAACAGAG GCAACAGAG aagtagtcto	GGAAGTCATC cgtgattcag gaggaatt tcggatgcatg ACCAGCCAG AACCAGCCAG ATACCCCCCCCCC CACCCCCCCCAC caccoctcgg aaactcataa ttcagaacc CTGCACTGG CTGCACTGG TCACACTACA GGCCCACTG GCCCACTGC tcaattcct tcaattcot tcaattcot tcattcot tcaattcot tcattcot tcaattcot tcattcot tcaattcot tcattcot tcattcot tcaattcot tcattcot tcattcot tcattcot tcaattcot tcttttaaattcot tcttttaaattcot tcattc
cccccacGGTGA gcccgggttc ggccactgag ggccactgag ggctacttc anartectga acactgatt gctacttca acacctrccc arggcagtga gggtgaactga ArccarGgca atagcatct AccTGCAAAA ACGCAGTTCAC CGACGTTCAC cacctcacaAAA	cacagGTTCC AGGGAGACCT AGGGAGACCT tctctgacc cactattcac ggtaggtct agcagcaccc GGCACTTCAG TGTGAGCAGG TGTGAGCAGG gttggagaa gttgtgagaa gttgtgagaa gttgtgagaa gttgtgagaa gttgtgagaa gtcgcacca ccgcaccac ACCTGGACTG ACCTGACCTG ACCTTCAGAT AGAGGACCAG aggcccctag agtatcccaa	TGCACTTGGC GGCCTTCCTC tgggacgtag cctoacott ttatgaccoa accaggaga GGAGCTGGCT GCTGCCTGGC GCTGCCTGGC tatgtgtgca ttaggtgtca GGTGCACTCT ggtcaggct ttggtgtcag GTGTCACTCC GGTGCCAGAG gtggtcgt agctgctcct gatggtcct gatggtcct	CATCATCCAT ANCTTCCAGA ACTTCCAGA ctgatgtagc tacttcaaat catgaggtca tagATCCCAC GTGATCCTGA AATCGTATGA ggccaacta aagtgcotat aagtgcotat tattgttett tTTGGTGTCCT cctogtgctt GAGCCCTGTA TGTTGAAGTG gaactactt gasctactt tattgtscat	CAAGAAAGG ACAACCTGCA agagtcacoc cagcccacat TCCACTTGC CTCCGAGCG GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC TGGGTGCTGA atgttgtgag ATGGCCGGAC TGGGGCTGAT AGGGGGCTGAT AGGCGGACTA AGGAGAGCTA gggctctacat ttttttttat	CACTGACCAT GCAGgtgege cagatottt cctagacgt GTGATCACC TTTCCAGGAA AGTCCTGCAC gcatgcccg gcagagctt tataagtc agaatagtg GGTGTCTACA TGTCAATGCT Cagaaattoc TGCCCTTCAC GTCAACAGAG TGACACAGAG ctaatgaggt agtatactt	GGAAGTCATC cgottgoctg ctgattcag gaggaatt ccaccccct ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC accocctcgg aactoctag attcagaacc aaggagtga CTTACCTCCT CAGgttggtg aatgcagcc CTCCCACTGC TTACCTCCT GGCCCACGTG TCACACTCA gcocttcct tcaatttct tttttaag
ctcttecc cGCCAGGTGA gcccgggtc ggccactgag ggctacttg ggctacttc AAATGCTGA ACACCTGCC CTCCACTGCC aggcgaggtg ggttgaaac attcgag attcgag attcatgoac taagcagtt atcCATGCT attactgoac taagcatt ACCTGCAAA GGCGTATTCT taactaCAAA CGGAGTTCAC taacctcaAA aggcattaac agagatta	cacagGTTCC AGGGAGACCT AGGGAGACC catattcac ggtgggtct agcaggacct totggtott GGCACTTCTG TGTGAGCAGG agtgggaga gttgtagaaaa gttgtggagaa gttgtgagaaa ACCTGGGCAT CCGGAGGCCA CCTTACCAGC CCTACCAGC ACCTTCAGAT AGAGGACGAG aggtacctag agtatccaa aaggcaact	TGCACTTGGC GGCCTTCCTC tgggacgtag cotaacott ttatgaccaa accaggaga cttgoatcg GGACCTGGCT GCTGCCTGGC GCTGCCTGGC CTGCAACTAC gggcaggtg ttcagaagat acttttttc CGTGGAGCTT gatcaggot ttggtcagg gTGTCACTCC GCTCCCAGGG gtgggtctgt agtcgctcct gatgtacct cotcccagact	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tagttcacat cacggagct tagACTCCAC GTGATCCTGA CAGCGTGGA AATGgtatgt ggccactta aagtgctat ttttgttctt tttgttctt GAGCCCTGGA CAGCCCAGC CCCCCAAGC CCGCCCAAGC gaactctt tatagaggg gcccccccc tatagaggt gccccccccc cagtagett	CAAGAAAGG ACAACCTGCA agagtcacca cagcccacat ttgcatttg ggggaactog TCCACTTGGC GCTCCGACC GTCCTGACTC dtgatgccat gtgatgccat gtgatgccat gtaccataa ccagCCACA AtGGCGCGAA AtGGCCGGAA ACCCCGATAC ACGCCGGACA gggcgctcaa gtactacat ttttttaat ttatatttgg gcgaqactc	CACTGACCAT GCAGgtogge cagatodtt octgagagt TGTGATCACC TTTCGAGAA AGTCCTGCAC ggectgecc gcagagett tataagtg GGTGTCTACA TGTCAATGC TGCCATCACG GCCACCACAGA AGCAGCACCAC CTGCACACAGA datagagtt aagtagtct gctatogaga	GGAAGTCATC ctgattcatg ctgattcatg cacccccct ttgqtccatg AACCAGCCAG ATACCCCCCC CACCCCCCCCC CACCCCCCCCCC
ctcttecc cccCaGGTGA gcccgCgtt ggccactgag ggctactgg ggctactto AAATGCTGA ACACCTTGCC CTCCACTGCC CTCCACTGCC cattggag gggtgaactgg gggtgaact ACCCATGGCA ACCCATGGCA ACCTGCAAA ACGGCATTTCT ACACTGCAAA CGGGAGTTCAC gggataacc gggataacc	cacagGTTCC AGGAGACCT AGGAGACCT agcaggtact agcaggtact GCACTTCTG TGTGAGCAGG TCCTGAAGGC actfggagaa gttggagaa gttgtgagaa gttgtgagaa gttgtgagaa gtcgcacact CCTGGCATT CCGGACTG CCCTACCAGC ACCTTCACATT AGAGGACCAG AGGCACAG aggcccctag agtaccacat ccatactgc	TGCACTTGGC GGCCTTCCTC tgggacgtag cctoacott ttatgaccoa accaggaga cttgattog GGAGCTGGCT GCTGCCTGGC TACCAACTAC ggggacgtg tatgtgtoa ttagtgtoa ttagtgtoag GTGTCACTCC CGTGGAGCTT gattaggct ttggtgtoag gTGTCACTCC GCGCCAGAG gtggtctat gatgtcact cactagct	CATCATCCAT ANCTTCCAGA attagattcacat catgatgtagc tacttcacat catggatgac tagAtTCCAC GTGATCCTAC CAGCGTGGGA AATCGtatgt ggccactta aagtgoctat aagtgoctat tattgttctt TTGGTGTCCT catogtgott GAGCCCTGTA TGTTGAAGTG gaactecttc gaactecttc tatcagaggg gcccccccacac cagtgagtt acggaggtca	CAAGAAAGG AcAACCTGCA agagtcacoc cagcccacat TCCACTTGC CTCCGACC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC AggagtCcag ATGGCCGCAC TGGGTGCTGA ATGGCCGCAC TGGGGCTGAT AGGCCGAC AGGACGCTA AGGCCGATA AGGCCGATA AGGCCGATA AGGCCGATA Ggcgtctacat tttttttatg ggaggagctc GTGTTTGGGAC	CACTGACCAT GCACGH2GC cagacotte ggtgtgggt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gcatgccoc gcasgactt tataagtc agaatagtc GGTGTCTACA TGTCAATGCT GACAACAGG GTCAACAGAG ctaatgagg catagtgg datagggt cgctaccdca gcatgggt	GGAAGTCATC cogottgootg ctgaattcag gaggaaatt ccaccoccac ttggtccatg AACCAGCCAG ATACCCCCCT CACCCCGCAC caccocctcgg aactoctag accocctcgg aactoctag cTTASCCTCT CAGgttggtg aatgcagcc CTCCCACTGC TCACGTCCA gcoattcoct ttottagact gcattcoct tttttaga gcoattcoct tttttag gcoattcoct tttttag gcoattcoct tttttag gcoattcoct
ctcttecc cGCCAGGTGA gcccggttc ggccactgag gdtacttg dtaggag gdtacttc AAATTGCTGA ACACCTTGCC CTCCATTCCA agggcagtg ggttgaac ATCCATGGCT attaggag tataggagtt ACCTGCAAA AGGCTATTCT ACACTAGAAA CGGAGTTAA CGGAGTTAC caactocc gtgcatacc agagttattt tottttttcc	cacagGTTCC AGGGAGACCT AGGGAGACC catatteac ggtgggtct agesgcaccc totggtott GGCACTTCTG TGTGAGCAGG agtggagaa gttggagaa gttggagaa gttgtagaaa gttgggtat ACCTGGCAT CCGGACACCG CCTCAACACCA CCTCCACAT AGAGCCCAG aggcaccta cacatagtgg cctcgtttgc	TGCACTTGGC GGCCTTCCTC tgggacgtag cctoacett ttatgacca accaggag dtgoattog GGACCTGGCT GCTGCCTGGC tcaccact ttaggacagtg tatgtgtca ttaggagat ttaggacagtg ttggtcag GTGCACACTGGGG GTGCCACAG gtgatctgt agctgctct agctgctcct cctcacact cctcacact cctcacact gtgatcact cctcacacact gcacgtacct cctcacacact gcacgtacct cctcacacact	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tactgaggect cacggggect tagACTCCAC GTGATCCTGA CASCGTGGGA AATGgtatgt ggccaacta aagactactt ttttgttatt tTGGTGGAAGTG CCGCCCAGC CAGCCCAGC CCGCCAGCC tatcagaggg ggcccccccag taaagtca tTGATGACTGT	CAAGAAAAG ACAACCTGCA agagtcacca cagcactagt TCCACTTGCC GCTCCGAGAC GTCCTGACTC Ctgoctocct agagtccag tggtgcctt gtccataa ccagCCACA Atgttgtgag ATGGCGGCTGAT ACGCCGGATAC ACGCAGCTA ggcdctaa gtatctact ttttttat ttatttag gcgagacc GTGTTGGAG	CACTGACCAT GCAGgtgcgc cagatottt cctgagagt TGTGATCACC TTCGAGGAA AGTCCTGCAC gcatgccc gcagagagtt tataagtg GGTGTCTACA TGTCAATGCT GGCATCACAGAG AGCAACAGAG CTGCACAGAG taatgagt aggttactot gctatggag GGTGTACAGAG GCCACGGGTC	GGAAGTCATC ctgattcatg ctgattcatg cacccccct ttgqtccatg AACCAGCCAG ATACCCCCCC CACCCCGCAC CACCCCCCCCCC
ctcttecc cccCAGGTGA gcccgggtta ggccactgag gctacttg gctacttc AAATGCTGA ACACCTTGCC CTTCCATCGC aggcaggtg ggttgaac ATCCATGGCT atagcatt ACCCATGGCA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGGAA ACCCATGAA A	cacagGTTCC AGGAGACCT tetetgaccc cactattcac ggtgaggtct agcaggtcet GCACTTCTG TGTGAGCAGG TCCTGAAGC agttggagaa gttgtgagaa gttgtgagaa gttgtgagaa gtgtgggat ACCTGGGCAT ccacattcat TCCTGACTG GCCTACCASC ACCTTCAGAT AGAGGACGAG aggcocctag agtatccaa acgggaact acggaact aggaact ccatattgg gcttgtttgc	TGCACTTGGC GGCCTTCCTC ttggacqtag cotcacottt ttatgacca accaggaga GGAGCTGGCT GCTGCCTGGC TACCAACTAC gggcaggtg ttatgtgca ttcgagagt ttggtgtcag GTGTCACTCC CGTGCAGCG gtggtctgt ggtggtctg ggggctgt gagtgctct gagtgtcag ctcacccacacc cocccagoct cocccagoct	CATCATCCAT ANCTTCCAGA anctTCCAGA ctgatgtagc tacttcaaat cacggagcat cacggagcat cacggagcat cacggagcat ggccaacta aagtgcotat aagtgcotat aagtgcotat aagtgcotat aagtgcotat cacggagct ttttgttett TTGGTGTCCT cctogtgctt GAGCCCTGTA TGTTGAAGTG gaactecttc tatcaagagg gccccccaac cggtagctt TGATGACTGT TATATTTGTA	CAAGAAAGG ACAACCTGCA agagtcacco cagcoccaat TCCACTTGC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC Aggagtccag ATGGCCGAC ATGGCGGCTGAT AGGCGGCTGAT AGGCGGCTGAT AGGCGGCTCAT ggcgtctcaa gtatctacat ttttttttat ttatattgg gcagagctc GTGTTGGGG CAAAAAAAA	CACTGACCAT GCCAGgtogc cagatottt cctagacgt GTCATCACC TTTCCAGCAA AGTCCTGCAC gcatgccc gcagagctt tataagttc agaatagtc GGTGTCTACA TGCCATCGT GTCAATGCT TGACCAATGCT GTCAACAGAG ctaatgagt catagagt aatgagtc agtatact gctatgcag gctaccgag gccagcag gcca	GGAAGTCATC cogottgootg ctgaattcag gagggaatt coaccccct ttggtcoatg AACCAGCCAG ATACCAGCCAG ATACCCCCTT CACCCCGAC aacoccotogy aactoctaa tttcagaacc aaggagtga CTTACCTCCA GGCCACGTTT CACGttgtg gocattcoct tttttaagg cocattcoct tttttaagg gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct ttttttaag gocattcoct tttttaagtcoa gocattcoct
ctcttecc cGCCAGGTGA gcccgGgtc ggccactgag ggctacttg dtaggag ggttactto AAATTGCTGA ACACCTTGCC CTCCACTTGCC ctrccACTGCC agggctgtg aggttgaag cattggag cattggag cattggag cattggag cattggag cattggag cattggag cattggag cattggag catcactgoc dtcaccocc dtagcotca ACCTGCAAAA ACCTAGAT ACCTGCAAAA CGGAGGTCAC gggcataac agagttattt tottttttec GTCCAAAGG GAAAAAAGG	cacagGTTCC AGGGAGACCT AGGGAGACCT agcaggcaccc totggtott GGCACTTCTG TGTGAGCAGG TCCTGAAGGC agttggagaa gttggagaa gttggagaa ACCTGGGCAT CCGGACTG CCTAACAGC CCTAACAGC ACCTTCACAGT AGCGCCCAG aggcacctag agtacccaa aggcaact ccaqtagtgg cctGAAAGAA	TGCACTTGGC GGCCTTCCTC tgggacgtag ctcaacttt gGAGCTGGCT GCTGCCTGGC GGAGCTGGCT GCTGCCTGGC CTGCAACTAC gggacagtg tatgtgtag attcagagat ttcagagat ttggtgtcag GTGCCACAG gtgatcgt gagtgtcdt gagtgacct gatgtacct gatgtacct agctgctcat gatgtacct cctccacatc agctgctcat gatgtacct cctcacact agctgctcat gagtgacct cctcacact agctgctcat gagtgtact	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tactgaggest cacgggtga tagACTCCAC GTGATCCTGA CACCCTGGA AATGgtatgt ggccaacta aagtoctat aagtoctat aagtoctat attgttatt tttgttatt tttgttatt tttgttagt cACCCCTGFA TGTTGAAGTG CCCCCAAGC CCCCCAAGC aagtocat accagtagtt tactagaggg gccccccac tacagaggt gccccccac cagtagatt TGATGATCTT TATATTGTA	CAAGAAAAGG ACAACCTGCA agagtcacca cagcactast ggggactcag TCCACTTGCC GTCCGAGAC GTCCTGACTC ctgoctocct aggagtccag tggtgcatt gtcataaa ccagCCACA TGGGTGCTGA TGGGTGCTGA ATGGCCGGAC TGGGGCTGAT ACCCGGATAC TGGGCCGAC TGGGCCGAC TGGGCCGAC TGGGCCGAC TGGGCCGAC GTGTTGGAG GCAAAAAAAA CACTGCACAC	CACTGACCAT GCAGgtgcgc cagatottt cctgagagt TGTGATCACC TTTCGAGGAA AGTCCTGCAC gcctgccc gcagagctt tataagttc agatagtg CGTGTCTACA TGCCATCAC GTCAACAGAG AGCAACAGAG CTCAACAGAG ctatgagt aggttactot gcttactot gcttactggaa GCGTTACTATTT TGCCCTACCCC GTTTAATTTT TGCCCTACCCC	GGAAGTCATC cogottgoctg ctgattcag gaggaatt ccacccccct ttgqtccatg AACCAGCCAG ATACCCCCCC CACCCCGCAC cacccctcg aactoctag attcagaacc aaggagtga CTTAGCTCT CACGttggtg aatgcagcc CTCCCACATGG TTACCTACCA GGGCCACTG GTCCCACTGC TCACAGTTCA gcattcct tttttaag atggagtca aggatacg gCTCTATAGACT
ctcttecc cccCAGGTGA gcccgCgtta ggccactgag gctacttg gctactta AAATTGCTGA ACACCTTGCC CTCCACTCCA aggcaggtg aggctgaact ATCCAATGCC ACACTGCAAAGGG GGCTATTCT ACACTAGAAA ACCCTGCAAAGG ggagataact tacttatagaact agaagcatta agaagcatta agaagtattt totttttec GTCCAAAGGG GTAAAAGAA ATTGTGCTAG	cacagGTTCC AGGGAGACCT tetetgaccc cactattcac ggtgaggtct agcagcaccc GCACTTCTG GCACTTCTG TCTGAAGGC agttgagaa gttgagaaa gttgtgagaa gttgtgagaa gggcoctgg aggcoctgg aggcoctgg aggcoctgg aggcoctgg ggtatccoa aggcaact ccafactgg GCTAAAGAA GAAAAATTT GATCACCCTC	TGCACTTGGC GGCCTTCCTC tdggacgtag octoacttt tatgacca accaggaga GGAGCTGGCT GCTGCCTGGC TACCAACTAC gggcaggtg tatgtgtoa ttaggtgtoa ttaggtgtoa ttggtgtoa GGTGCCCCCAGG GGCGCCCTA Gagtagtact agctgctoct gagtagtact coctcagoct coctcagoct coctcagoct CACCTGGGCTGA AAAGGTGTA	chitchtcarcan Anctrocada cigatgitago tacttcaaat catgaggitago tagAttcacat dagAttcac tagAttcac tagAttcac ggocactta aagtgoctat aagtgoctat aagtgoctat tittgitett TTGGTGGCCT cotogigott GAGCCTGTA AGTGAGATG gaactoctto tatcagaggg ggococococo gtaagtcat catagaggg ggococococo gtaagtcat CatacATTGTA CTTATATTGAA CTTATATCA	CAAGAAAGG ACAACCTGCA agagtcacoc cagcecaat TCCACTGCC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC TGGGTGCTGA ATGGCCGAC ATGGCGGCCAA ATGGCCGGAC TGGGGCTGAT ACCCCGACCAC GTGTTGGAG CAAAAAAAA CACTCGACC CTTTTGAG CATTGACC	CACTGACCAT GCACGUGGC Cagacotte ggtgtgggt TGTGATCACC TTTCCACGAA AGTCCTGCAC gccatgccoc gccatgccoc gcagagctt tataagtte agaatagte GGTGTCTACA TGTCAATGCT GGCAATGCT TGACCACAGAG ctaatgagt aatgagtte aggtatgaga gctatgaga gctaccgra GTCTTAATTT TGCCTACCCC GGCCACAAA	GGAGGAGTGATC orgottgootg ctgaattcag gaaggaatt ccacccccct ttggtccatg AACCAGCCAG ATACCCCCT CACCCCGCAC acoccotogy aactoctaa tttcagaacc aaggagtg CTTASCCTC CAGttgtgt GGCCAGTGC TCACAGTCA gcoattcoct tcaattct tttttaagg gcasttcoct tcaattct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tttttaag gcasttcoct tcaactTTG ACCCTTGA
cccccacget gcccacggtc ggccactgag ggctactgg ggtacattg gctactto AAATTGCTGA ACACCTTGCC CTTCCATTCCA agggcagtg gggttgaaac cattggaag cagactgtg gggttgaaac ATCCATGGCT ataagcatct ACCTGCAAAA GGGCTATTCT ctcacctccc gagactata CGGCGACACCACAAA GGGCTATTCT gggataacc gggataacc gggataacc gggataaca gagtttttttte cttttttec GAAAAAAGA ATTGTGGTAG	cacagGTTCC AGGGAGACCT AGGGAGACCT agesgcaccc totgttott GCCACTCTG TGTGAGCAGG TCCTGAAGGC acttgsgcac gttgsgcac gttgsgcac tccsgcat tccsgcat CCCTGACGGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC aggcaccta aggcaccta aggcaccta ccastafytg cctGAACAGC GCTGAACAGC	TGCACTTGGC GGCCTTCCTC tgggactag ctcaacttt tatgaccaa accagagag ctgactog ctcacttg ctcactact gGAGCTGCCTG CGCCCTGGC CGTGCACTAC agtgacagtt ttogtgtcag GTGTCACTAC GCTGCCACAG gtgagtcgt cctcCCACAG gtagtgtcat cotcacact agctgctcat gatgtact cotcacact AAGGCTGA AAGGCTGA AAGGCTGA	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tactgaggest cacggggtga tagACTCCAC GTGATCCTGA CACCGTGGA AATGgtatgt ggccaacta aagtgctatt ttttgttctt tTGGTGGACTGT GAGCCCTGTA TGTTGAAGTG GAGCCCTGTA TGTTGAAGTG gaactactt taccaggggt gccccccacc gtaaagtca ccgtaagttca CGGTGGGGG CTTTTGTGGGG	CAAGAAAAGG ACAACCTGCA aggtcacce cagcccacat TCCACTTGCC GTCCGAGAC GTCCGAGAC GTCCGAGAC GTCCGACC aggatccag tggatgcctt gtccataaa ccagCCACA ATGGCGCCGA TGGGTGCTGA ATGGCCGGAC TGGGGCTGAT ACCCGGATAC GGGCCCGAT ACCCGGATAC GGGCCCGAT ACCCGGATAC GGCCCGAT CCCGGATAC CTTTTTGGAG CAAAAAAAA CACTCCACAC CTTTTTGGAGA	CACTGACCAT GCAGgtggg GCAGgtggg Cagatodtt octagagagt TGTGATCACC TTTCGAGGAA AGTCCTCGAC ggcdtgccco gcaagagtt tataaagtc agataggtg CGTGTCTCAC TGCACTTCAC GTCAACAGAG AGCAGCAGCA GCAACAGAG ctatgagt gcttactot gcttaggag GCGTTACCAGG GCAACGACG GCAACGAGA GCTAGCCCG GGGACGACAA AGGTTATCAA	GGAAGTCATC cogottgoctg ctgattcag gaggaatt ccacccccct ttggtccatg AACCAGCCAG ATACCCCCCC CACCCCGCAC CACCCCGCAC acocctcgg aactoctaa ttcagaacc aaggagtga CTTAGCTCT CACGttggtg aatgcagcc CTCCCACATGG TTACCTACCA GGCCACTG goctaccGCT ttatttaag aggagtca atggagtca atggagtca atggagtca atggagtca ACCTATCA AAATTCATCA
ctcttecc cccCAGGTGA gcccgggttc ggccactgag ggctacttg gctacttc AAATTGCTGA ACACCTTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCA AGGCTATTCT ACACTAGAAA CTGGCATATCC CGGAGTTCAC CGGAGTCAC taccctcc ggcataacc ggagataac gagagatta agagttttt cttttttco GTCAAAAGGA AATTGTGGTA AATTGTGGTA AATTGTGGTA	cacagGTTCC AGGGAGACCT tetetgaccc cactattcac ggtgaggtct agcaggtcc GCACTTCTG GTGAGCACG TCCTGAAGGC agttgagaa gttgtgagaa gttgtgagaa gttgtgagaa gtgtgtat ACCTGGGCAT cogacgcat CCCTACCASC ACCTTCAGT aggcocctag aggcocctag aggcocctag aggcocctag catgtttgc GCGAAGAAATTT GAACAGACCCT AAACTTCTTT	TGCACTTGGC GGCCTTCCTC tgggacgtag cotoacott ttatgaccoa accaggaga cttgattog GGACCTGGC TACCAACTAC gggcaggtg ttatgtgca ttagtgtca ttggtgcagt ttggtgcag GGTGCACTGC GCTGGCGGGG GCTGCCCAGGG gtgagtctgt gagtagtact coctcagoct coctcagoct coctcagoct CACCTGGGGTGTA AAAGGGTGTA AAAGCTGACTG	CATCATCAAT ANCTTCCAGA ctgatgtagc tacttcaaat catggagcat catggagcat catggactat tagACTCCAC ATGGTGGGA AATGgtatgt ggccaacta aagtgoctat aagtgoctat aagtgoctat catggtgott ttttgttett TGGTGGTCCT cccccCAASC AGTGAGATG gaactacttc tatagaggg gcccoccoac gtaagtcat ccgqtagtt Gaactacttc tGATGACTGT TATATTGTAA CTTATGGAA CTTATGGAAC	CAAGAAAGG ACAACCTGCA agagtcaocc cagccacaat TCCACTTGC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC TGGGTGCTGA AGGACGCTGAT ACCCGGAC TGGGCTGAT ACCCGGATC AGGACGGATA AGGACGGATA AGGACGGATA AGGACGGATC GTGTTGGAG CAAAAAAAA CACTCCACACC CTTTTGGAG	CACTGACCAT GCCAGtgagagt cagacotte ggtgtggt TGTGATCACC TTTCCAGGAA AGTCCTGCAC gccatgcco gcagagagt tataagtte agatagtte GGTGTCTACA TGTCATCGT GGCCTTCAC GTCAACGAG GTCAACGAG GTCAACGAG agettacta gctatgaga cgtccggaa cgctccggaa GGCACGGTCT GTTTTATTTT TGCCTAGCCC GGGACGAGAA AGGTTATCAA	GGAAGTCATC ogottgootg ctgaattcag ccaccccct ttggtccatg AACCAGCCAG AACCAGCCAG AACCAGCCAG aacoccccg aactoctaa tttcagaacc aaggagtg CTTAGCCTCT CAGGttgtg GGGCCAGTG TTACCTACG GGGCCAGCTG TCACAGTTCA GGGCCAGCTG tttttaag atggatcaa gctaactg tttttaag tttttaag tttttaag AAAAGTCTTG AAAACTTTG AAAACTTTG AAAACTTTG AAAACTTTG AAAACTTTG AAAACTTTG AAAACTTTG
ccccaccactaga gccccgggttc ggccactgag ggccactgag ggctacttg ggtaggtg agatagtg ggttactto AAATGCTGCC CTCCACTCCA agggcagtg gggttgaaac ATCCATGGCT ataagcatgt ACCCTGCAAAA GGGCTATTCT ACCATAGAT CGAGAGTCAC agagtatact tottttttc grgcataacc grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg grgcataacg gracaacg agagttattt tottttttc attotfaga ATTCTGGTAG AATTCTGAGA	cacagGTTCC AGGGAGACCT AGGGAGACCT cottgagcaccc cotattcac GCCACTCTG TGTGAGCAGG TGTGAGCAGG gttgagcac gttgagcac gttgagcac gttgagcac gttgagcac dtottaaa gttgtgagta ACCTGGCAT ccacattcatt TCCTGACGG CCCTACCAGC ACCTTCACAGA AGCTGCACGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACCAGC CCCTACAGC GCTGAACAG GAAAAAATT GATCAGCCCT AAACTTCTTT ATACGCGTCA	TGCACTTGGC GGCCTTCCTC tgggactag cctacactt tatgacca accagaga GGAGCTGGCT GCTGCCTGGC GCTGCCTGGC CGTGCAGCT gggcagtg tatgtgca ttcgtgtcag GTGTCACTCC TCACCTGGGG gtgagtcgt ccTGCCACAG gtgagtcgt cctoccacac agctgctcat gatgtcct catcagact catcacactTG AAAGGTTGA CATTGTCGTTG TAAACCTCAC	CATCATCCAT ANCTTCCAGA ANCTTCCAGA ctgatgtagc tactgaggest cacggggtga tagACTCCAC GTGATCCTGA CAGCGTGGA AATGgtatgt ggccaacta aagtgoctat aagtgoctat aagtgoctat TGGTGGTCCT cctogtgott tTTGGTGCTCT CatCagtgott tatcaaggg ggcccccacac gtaaagttca coggtagatt TGATGACTGT TATATTTGTA CTTATATCCA CTTTTGTGGGG CGACATGACTG	CAAGAAAAGG ACAACCTGCA aggtcacca cagcccacat TCCACTTGCC GTCCGAGAC GTCCGAGAC GTCCGAGAC GTCCGAGAC GTCCGAGAC TGGGTGCTGA TGGGTGCTGA Atgtcgag ATGGCCGAC TGGGGCTGAT ACCCGATAC GGGCTGAT ACCCGATAC GTGTTTGGAG CAAAAAAAA CACTGCACCG TTTTTGGAGA ACCCCCGTGC	CACTGACCAT GCAGgtggg GCAGgtggg GCAGgtgggt GCAGgtgggt GCAGgtgggt GTGTATCACC TTTCGAGGAA AGTCCTGCAC GGAGAGG GCAGGCC GCAGTCTCAC GCAGTCACAGAG TGCCATCAC GCAACAGAG AGCAGCAGCT GCCACTCAC GCAACAGAG GCAACAGAG GCAACAGAG GCAACAGAG GCAACAGAG GCAACAGAG GCAACAGAG GCAACAGAG GTTTATATTT GGCGACGACAA AGGTTATCAA TATCCTGTGA	GGAAGTCATC cogottgoctg ctgattcag gaggaatt ccacccccct ttggtccatg AACCAGCCAG ATACCAGCCAG ATACCCGCCAC CACCCCGCAC CACCCCGCAC CTTAGCCTCT CAGgttggtg aatgcagcc CTCGCACATGG TTACCACCA GGCCAAGTCA GGCCAAGTCA GGCCAAGTCA gcottaccttc tttttaag atggagtcoa gottaacgtg GACGTTATGA AAATTCATG CATCTACAA AAACGTCTT AAATTCATG CATCTACAAG
ctcttecc ccccAcGTGA gcccgggttc ggccactgag ggctacttg gctacttc AAATTGCTGA ACACCTTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC CTCCACTGCC agggctgaac ATCCATGCCT acattggag gggttgaac ATCCATGGCA acattggac agagctata acagctct ggcataacc gggataacc gggataacc ggcacaacta agagtttt tctttttc CTCCAAAGGG GTAAAAGAA ATTGTGGTAG AATTGTGGTAG	cacagGTTCC AGGGAGACT tetetgaccc cactattcac ggtgaggtct agcaggtct GGCACTTCTG GTGAGCACG TCCTGAAGGC agttggaaa gttgtgaaaa gttgtgaaaa gttgtgaaaa gttgtata accTGGGCAT ccgacgcact ccgacgCact ccGCTACCAGC ACCTTCAGAT AGAGCACCAG aggcoctag agtateccaa aggtagttt ccGAAGAAAATT GATCAGCCCT AAACTTCTTT TACAGCGTCA	TGCACTTGGC GGCCTTCCTC tgggacgtag cotoacott ttatgaccoa accaggaga GAGCTGGCT GCTGCCTGGC TACCAACTAC gggcaggtg ttatgtgtca ttagtgtcag GTGTCACTCC GGGGGGCTT gadtcaggot ttggtgtcag gtgatctgt agtgtact gatgtact gatgtact coctcagoct coctcagoct coctcagoct CAATTTGTGTG TAAACCTCAC GTATTTGGTG GTAGTGGGA	CATCATCAAT ANCTTCCAAG ctgatgtagc tacttcaat catggagct tagACTCCAC GTGATCCTAA AATGgtatgt ggccactta aagtgcotat aagtgcotat aagtgcotat aagtgcotat aagtgcotat ctotggtt GAGCCCTGTA TGTTGAAGTG GCCCCAAGC AGTGAGATG gactccttc tactaggagt gccoccccc gtaagtcactta CCGCCCAAGC AGTGAGATG GTGAAAGTG CTTTTGTGGAG CTTTTGTGGAG CGACTGACTG GTGAAAAGTT	CAAGAAAGG ACAACCTGCA agagtcaccc cagcccaat TCCACTGCC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC GTCCTGACTC CTGGGCCGAA TGGGTGCTGA ACCCGGCTGAT ACCCGGAC TGGGCCGAA ACCCCGACA GGGCTGAT ACCCGGACA CTTTTGGAG CAAAAAAAA ACCCCACCC CTTTTGGAG ACCCCACTGG ACTACCTGAC	CACTGACCAT GCCAGtgogo cagatottt octgagodto ggtgtgggt GTGTATCACC TTTCCAGGAA AGTCCTGCAC gcatgoco gcagagott tataagtto agatagtg GGTGTCTACA TGTCATCGT GGCATCACAGAG GTCAACAGAG GTCAACAGAG GCCACCACAGAG gctatgagt aggttactt gctatgag gccASCGTCT GGCTACCC GGACGACAA AGGTTATCAA	GGAGTGATC cogottgoctg ctgaattcag ccacccccct ttggtccatg AACCAGCCAG AACCAGCCAG AACCAGCCAG aactoctag aactoctag cttccagaacc aaggagtg aatgoagcc CTGGCAGTGG TTACCTACCA GGGCCAGCTG TCACGAGTCA gocattcact tttttaag atgggtca gottaactt tttttaag atgggtca gACTTATAAAA AAAACGTCTT AGATCATTGA AAATTCATGA AATTCATGA ATTAAAGGA

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- κ B sites and the transcription start site have been previously characterized (Le Bail *et al.*, 1993; Ito *et al.*, 1994). The three NF- κ B sites in the promoter and the one site in the first intron are in boldface type and are underlined. The transcription 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 $I\kappa B\alpha$ gene family, a phylogenetic tree was constructed based on the Ankyrin repeats of protein sequences of a variety of members of the $I\kappa B\alpha$ 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 I κ B α 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 I κ B α 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 I κ B α 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- κ B, respectively. The I κ B α 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- κ B1 and 2. These proteins associate with NF- κ B1 (p50) or NF- κ B2 (p52) and seem to be grouped accordingly. This grouping suggests that the Ankyrin repeats present in NF- κ B1 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- κ B1 and 2 and is placed in a separate group. However, the branch between Bcl-3 and NF- κ B1 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 $I\kappa B\alpha$ 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- κ B2. 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- κ B1 and NF- κ B2 in one clade and Dorsal, c-rel, and RelA (p65) on another (Schmid *et al.*, 1991). Likewise, the I κ B α gene family also diverges into a group that interacts with either or both NF- κ B1 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

$\begin{array}{c ccccccccccccccccccccccccccccccccccc$														
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		1 2	2 3	4	5	6	7	8	9	10	11	12	13	14
11 1.0000 0.9404 0.8468 0 12 1,0000 0.8426 0	1 2 3 4 5 6 7 8 9 10 11 12	1.0000 0.82 1.00	201 0.3718 200 0.3419 1.0000	4 0.3621 0.3319 0.9009 1.0000	0.3609 0.3640 0.5299 0.5431 1.0000	0.3722 0.3808 0.5299 0.5388 0.9560 1.0000	0.3609 0.3682 0.5043 0.5259 0.8388 0.8498 1.0000	0.3722 0.3766 0.5940 0.5560 0.5421 0.5438 0.5458 1.0000	9 0.3797 0.3975 0.5214 0.5345 0.6044 0.6022 0.6007 0.6751 1.0000	0.4000 0.3787 0.4316 0.4483 0.5191 0.5277 0.5106 0.4766 0.5234 1.0000	0.3915 0.3574 0.4231 0.4483 0.5149 0.5191 0.5021 0.4723 0.5106 0.9617 1.0000	$\begin{array}{c} 12\\ 0.3872\\ 0.3617\\ 0.4316\\ 0.5106\\ 0.5191\\ 0.5021\\ 0.4809\\ 0.5021\\ 0.9532\\ 0.9404\\ 1,0000\\ \end{array}$	0.3830 0.3532 0.4103 0.4181 0.4851 0.4979 0.4979 0.4979 0.4596 0.4979 0.8553 0.8468 0.8426	14 0.3410 0.3682 0.4444 0.4569 0.4291 0.4291 0.4406 0.4674 0.4598 0.5234 0.5234 0.5191

Note. 1, *Drosophila* Notch; 2, human Tan-1; 3, human Bcl-3; 4, mouse Bcl-3; 5, mouse NF- κ B1; 6, human NF- κ B1; 7, chicken NF- κ B1; 8, human NF- κ B2; 9, chicken NF- κ B2; 10, pig I κ B- α ; 11, human I κ B- α ; 12, rat I κ B- α ; 13, chicken I κ B- α ; 14, *Drosophila* Cactus.





FIG. 4. Phylogenetic analysis of the $I\kappa B\alpha$ 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 $I\kappa B\alpha$ 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- κ B sites have been identified in the porcine, murine, and human $I\kappa B\alpha$ 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 I κ B α gene also contains a putative NF- κ B 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 $I\kappa B\alpha$ promoter is relatively weak (personal observations) and may require an enhancer to augment its activity. NF- κ B and $I\kappa B\alpha$ have been linked in an autoregulatory loop (reviewed in Beg and Baldwin, 1993), and this intronic NF- κ B site may also participate in the regulation of the I κ B α gene. The intron sequences of the I κ B α gene in other species or other family members is not available for comparison. The finding of this NF- κ B site in the introns of the I κ B α 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). I κ B: A specific inhibitor of the NF- κ B 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 λ gt 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- κ B 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 I κ B α 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 I κ B α -like gene is regulated by NF- κ B. *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 remov-

ing 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 IkB 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 Relparticipants 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 I_KB-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- κ B sites in the I κ B α promoter are required for induction of gene expression by TNF α . *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 $I\kappa B\alpha/MAD3$ 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- κ B and I κ B/MAD-3 inhibitor of the NF- κ B to 4q24 and 14q13, respectively. *Genomics* **14**: 529–531.
- Mathew, S., Murty, V. V. S., Dalla-Favera, R., and Chaganti, R. S. K. (1993). Chromosomal localization of genes encoding the transcription factors c-rel, NF- κ B p50, NF- κ B 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- κ B 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- κ B precursor p105 and the proto-oncogene product Bcl-3 are I κ B molecules and control nuclear translocation of NF- κ B. *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 UVinduced 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- κ B 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.