Cell, Vol. 62, 1019–1029, September 7, 1990, Copyright © 1990 by Cell Press

Cloning of the p50 DNA Binding Subunit of NF-κB: Homology to *rel* and *dorsal*

Sankar Ghosh,* Ann M. Gifford,* Lise R. Riviere,† Paul Tempst,† Garry P. Nolan,* and David Baltimore*‡
* Whitehead Institute for Biomedical Research
Nine Cambridge Center
Cambridge, Massachusetts 02142
† Howard Hughes Medical Institute
and Department of Genetics
Harvard Medical School
Boston, Massachusetts 02215
‡ Rockefeller University
New York, New York 10021

Summary

The DNA binding subunit of the transcription factor NF-κB, p50, has been cloned. p50 appears to be synthesized as a larger protein that is then processed to its functional size. Sequence analysis reveals remarkable homology for over 300 amino acids at the amino-terminal end to the oncogene v-*rel*, its cellular homolog c-*rel*, and the Drosophila maternal effect gene *dorsal*. This establishes NF-κB as a member of the *rel* family of proteins, all of which display nuclear–cytosolic translocation. Protein sequence from the p65 polypeptide has established that it is not encoded in the same mRNA as p50. However, p65 appears homologous to c-*rel*, suggesting that c-*rel* may form heterodimers with p50 and *rel* may include a homodimerization motif.

Introduction

The regulated expression of genes requires the binding of multiple sequence-specific DNA binding proteins to *cis*-acting promoter and enhancer elements (Mitchell and Tjian, 1989; Dynan, 1989). The tissue- and developmental stage-specific expression of different genes can be explained by the restricted presence of these regulatory proteins in various tissues. While most of these regulatory transcription factors are themselves controlled in their expression at the level of transcription, there are examples of factors that are present in all cells but active only in some cells because of posttranscriptional activation (Picard and Yamamoto, 1987; Larson et al., 1988). The best studied example of the latter kind is the immunoglobulin κ enhancer binding protein, nuclear factor κB (NF-κB; Sen and Baltimore, 1986a).

NF-κB was identified initially as a DNA binding protein, present only in the nucleus of mature B cells, that specifically complexes with the κB site in the immunoglobulin enhancer (Sen and Baltimore, 1986b). Since then, it has become evident that NF-κB is not only a regulator of gene action during development but is also involved in the inducible expression of a large number of genes in different cell types (for review see Lenardo and Baltimore, 1989; Baeuerle and Baltimore, 1990). These include cytokines, cytokine receptors, major histocompatibility antigens and associated proteins, serum amyloid A protein, and a variety of viruses including human immunodeficiency virus, SV40, and cytomegalovirus. Although NF-κB is by no means the sole element determining the inducible expression of these genes, it plays a significant role. A characteristic feature of NF-κB induction is that it occurs without synthesis of new protein. In its uninduced form, NF-κB is present in the cytosol bound to an inhibitor protein, IκB (Baeuerle and Baltimore, 1988a, 1988b). Upon induction by a variety of agents, including bacterial lipopolysaccharide, phorbol 12-myristate 13-acetate, tumor necrosis factor α, double-stranded RNA, and interleukin 1, the cytosolic NF-κB–IκB complex is dissociated and the free NF-κB translocates to the nucleus (Baeuerle and Baltimore, 1990). Recently it has been demonstrated that some of these induction signals work by activating cellular kinases that can then phosphorylate IκB (Ghosh and Baltimore, 1990). The phosphorylated IκB cannot associate with NF-κB, thereby resulting in the release of free NF-κB. Therefore the NF-κB–IκB system is a unique example of control over the function of a protein by cytoplasmic/nuclear partitioning: the NF-κB protein can be considered a second messenger, transmitting signals from the cell surface to the nucleus.

The wide role of NF-κB along with its unique mechanism of activation makes its regulation and function of particular interest. Its low abundance, however, has frustrated efforts at detailed characterization. Antibodies to NF-κB are not available, and numerous attempts to clone NF-κB using oligonucleotide binding to λgt11 fusion proteins have failed (Singh et al., 1988; Fan and Maniatis, 1989). We have therefore undertaken to purify sufficient protein for direct sequencing. Purified NF-κB consists of two polypeptides, of M_r 50,000 (p50) and 65,000 (p65) (Baeuerle and Baltimore, 1989; Ghosh and Baltimore, 1990). The p50 subunit binds to DNA, while the p65 subunit is necessary for NF-κB interaction with IκB. Using a combination of conventional chromatographic steps and oligonucleotide affinity chromatography, we have obtained large quantities of the purified protein from rabbit lung tissue. With primers deduced from protein sequence, a specific probe was synthesized by polymerase chain reaction (PCR) that was then used to obtain a full-length cDNA clone encoding the DNA binding subunit of NF-κB.

We report here that the p50 subunit of NF-κB is encoded by an mRNA containing an open reading frame of 107,000 daltons and therefore appears to be processed down to 50,000 daltons. DNA binding by this protein is carried out by a novel, as yet undefined, motif. The sequence of the amino-terminal half of the protein, which is the region that includes p50, is remarkably similar to the protooncogene *c-rel*, its viral oncogenic counterpart v-*rel*, and the Drosophila maternal effect gene *dorsal* (Steward, 1987). In addition, we have found that p65, which forms a

heterodimer with p50, is probably a novel member of the *rel* family of proteins and suggests a role for the *rel*-homologous regions as dimerization motifs.

Results

Purification of NF-κB

Purification of NF-κB was carried out from rabbit lungs, which contain large amounts of cytosolic NF-κB–IκB (Ghosh and Baltimore, 1990). Because large losses in yield were incurred through multiple rounds of chromatography on an affinity resin to which was bound an oligonucleotide containing the κB site, as described previously (Baeuerle and Baltimore, 1989; Ghosh and Baltimore, 1990), the purification scheme was modified by the incorporation of a negative selection step (Kawakami et al., 1988). Partially purified NF-κB from an S-Sepharose column was first bound to the κB site in the presence of competitor DNA, eluted at 0.4 M KCl, and then rebound to a mutant κB site (Kawakami et al., 1988) and efficiently eluted at 0.2 M KCl. Nonspecifically bound proteins still eluted at 0.4 M KCl. The 0.2 M KCl eluate was repurified on the wild-type κB site, and the final 0.4 M KCl fraction contained NF-κB in good yield and high purity. We obtained about 30 μg of NF-κB from 2 kg of tissue.

During the purification we observed a gradual decrease in the amount of the p65 polypeptide relative to p50, indicating a significant dissociation of the complex. The final purified fraction displayed the two major bands of p50 and p65 upon silver staining of an SDS–polyacrylamide gel (Figure 1A). To confirm that these major polypeptides were indeed the two subunits of NF-κB, they were eluted out of an SDS–polyacrylamide gel, renatured, and assayed for various properties of NF-κB that have been reported earlier (Baeuerle and Baltimore, 1989). As expected, the p50 polypeptide bound specifically to the κB site, formed a heterodimer with the p65 polypeptide, and was inhibited in its DNA binding by IκB (data not shown). Therefore we believe that the major polypeptides in the purified preparation are the subunits of NF-κB.

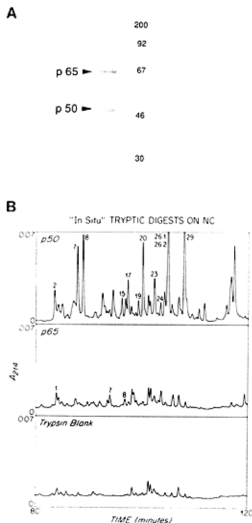


Figure 1. Tryptic Digestion of Purified NF-κB

(A) SDS–polyacrylamide gel electrophoresis of the final affinity-purified NF-κB preparation. Ten microliters of the purified protein was analyzed by electrophoresis on an 8.75% SDS–polyacrylamide gel. Proteins on the gel were stained with silver.

(B) Reverse-phase HPLC of tryptic peptides from the p50 and p65 subunits of NF-κB. Top, absorbance profile of peptides from p50; middle, absorbance profile of peptides from p65; bottom, profile of a trypsin blank (see Experimental Procedures). Numbered peaks in the top panels correspond to sequenced peptides in Table 1. HPLC conditions were as described in Experimental Procedures. Only relevant sections of the chromatograms are shown; full scale corresponds to 0.07 absorbance unit at 214 nm.

Sequencing the p50 and p65 Subunits of NF-κB

Purified NF-κB was fractionated by SDS–polyacrylamide gel electrophoresis, and the proteins were transferred to nitrocellulose (Aebersold et al., 1987). The 50 kd and 65 kd polypeptides transferred from the gel with strikingly different efficiencies, yielding only about 5% of p65 on the nitrocellulose membrane compared with about 50% for p50 (data not shown). Direct amino-terminal sequencing was attempted, but both termini appeared modified or blocked. The p50 protein on nitrocellulose was then digested with trypsin (Aebersold et al., 1987), and the resulting tryptic peptides were separated by high pressure liquid chromatography (HPLC) (Figure 1B). Ten purified peptides were sequenced, and a total of 166 amino acids of primary protein sequence was obtained (Table 1). (The sequence total is 178 amino acids if the arginine and lysine residues that precede the tryptic peptides are included.) In contrast, owing to the extremely small amounts of protein, sequencing of the p65 polypeptide resulted in only 40 amino acids of primary sequence from three peptides (the sequence from peptide 1 is shown in Table 1).

Molecular Cloning of the p50 DNA Binding Subunit of NF-κB

Oligonucleotide probes deduced from the first 7 amino acids of peptides T7.1 and T24 of p50 were used as primers for PCR-mediated amplification of the sequences flanked by the primers (Lee et al., 1988). cDNA made from total rabbit spleen and lung RNA was used as template. The degeneracy in the third position of the genetic code resulted in 256- and 2048-fold degenerate primers corre-

Table 1. Amino Acid Sequences of Tryptic Peptides from the p50 and p65 Subunits of NF-κB

Peptide Number	Initial Yield (pmol)	Sequence
T2a	5.0	NIHLHA
T2b	2.5	x(DI)QIR
T7.1	3.5	HCEDGICTVNAG(PK)
T8.1	10.2	YVCEGPSHGGLPGAS(S)EK
T15a	6.3	LYYPEIK
T15b	2.5	(V)IVQLVTDG
T17.2	19.0	RLEPVVSDAIYDSK
T19	4.0	PASVFVQLR
T20.1	22.0	TAGCVTGGEEIYLLCDK
T23.1	8.0	xFAIVF(K)
T24	3.0	(D)VNITKPASVFVQLR
T26.1	5.5	(S)DLETSEPKPFLYYPEIK
T26.2	6.5	xYNPGLLVHPDLAYLQTE(GGG)D
T29.7	11.7	FYEEEENGGI(W)EGFGDFSPT(D)VH
T1a(p65)	2.0	xxFSQADVHR

Sequences are in single-letter amino acid code. Residues in parentheses are tentative; "x" indicates that no positive identification could be made. Initial yields of all sequencing experiments are listed. Peptides were separated on a narrow-bore C8 column and are indicated on the top chromatogram shown in Figure 1B. Peptides T7, 8, 17, 20, 23, 26, and 29 were repurified on a C18 column. Peak fractions T2 and T15 from p50 and T1 from p65 contained two peptides.

sponding to peptides 7.1 and 24, respectively. Because the relative positions of the primers on the protein could not be determined, the primers were made in both orientations and were used in all combinations. Only primers 2 and 3 gave an amplified fragment; it was amplified from both lung and spleen cDNAs and was approximately 700 bp long. The fragment was cloned into a plasmid and upon sequencing predicted an open reading frame that contained four of the peptide sequences (data not shown), thereby proving that the amplified fragment was indeed derived from mRNA encoding the p50 polypeptide.

The length of the PCR-amplified fragment gave us the confidence to use this rabbit probe to screen mouse cDNA libraries, because previous biochemical characterization had indicated that the mouse and rabbit factors were very similar. When an oligo(dT)-primed cDNA library made from mouse 22D6 cells was screened with the PCR-amplified fragment, five positive clones were obtained. All of the clones contained an insert of approximately 4.0 kb, and they were shown to be identical by restriction enzyme mapping. One of the clones, clone 3, was chosen for further analysis and sequencing.

Sequence of the cDNA Encoding p50

The 4.0 kb cDNA was sequenced completely on both strands by using progressively shorter clones generated by exonuclease III deletion. Regions not covered by the deletion clones were sequenced using specific oligonucleotide primers based on newly acquired sequences (Figure 2 and Experimental Procedures). There was a long poly(A) tail at the 3′ end of the clone that was preceded by 3892 bp of heteropolymeric sequence. A long open reading frame encoding a predicted protein of 971 amino acids (M_r 107,000) was found, starting with an ATG codon at position 290 and terminating with a TAG codon at position 3204. The open reading frame is preceded by an in-frame termination codon at position 252, indicating that the ATG at 290, which satisfies the consensus rules of Kozak (1989), is probably the true initiating ATG. Examination of the frequency of rare-codon usage also indicates that the sequences preceding the ATG at 290 do not code for any protein. A putative polyadenylation signal sequence is present at an appropriate distance from the poly(A) tail.

The sequence of the PCR-amplified fragment was found in the 5′ portion of the clone between nucleotides 633 and 1278, and all of the peptide sequences obtained from direct protein sequencing were evident and are indicated in Figure 2. There were just 5 amino acid changes between the 180 amino acids in the rabbit peptides and the predicted mouse protein, 3 of which were conservative. This translates into 98% identity in this region of NF-κB between these divergent species. All of the peptide sequences were found in the 5′ half of the clone (Figure 2); therefore the p50 protein is probably confined to approximately the first 450 amino acids of the open reading frame, while the remainder potentially encodes an additional polypeptide of 60 kd. A highly glycine-rich segment around amino acid 400 may act as a signal for a processing event, suggesting a novel mechanism by which NF-κB, an intracellular cytosolic protein, is synthesized as a larger protein but undergoes a maturation step that processes it into a 50 kd polypeptide.

The predicted amino acid sequence of p50 NF-κB was used in a homology search of computer data bases and revealed striking homology, extending from amino acids 39 to 374, with the turkey proto-oncogene *c-rel* (Wilhelmsen et al., 1984), its corresponding viral oncogene *v-rel* (Stephens et al., 1983), the mouse and human *c-rel* (Grumont and Gerondakis, 1989; Brownell et al., 1989), and the Drosophila maternal effect gene *dorsal* (Steward, 1987) (Figures 3A and 3B). The region of the clone encoding p50 has about 45% identity with *v-rel*/*c-rel* and 50% identity with *dorsal*. If conservative amino acid changes are considered, the similarity increases to about 60% over this region. At the carboxyl terminus, the proteins diverge completely from one another beginning around position 376 of p50 NF-κB (Figures 3A and 3B), indicating that the regions homologous to *c-rel* perform a common function in all these proteins while the remainder of the molecules might have divergent functions (Steward, 1987).

The other *rel*-related proteins, like NF-κB, are localized at some stages to the cytosol and at other stages to the nucleus. In keeping with this role, a putative nuclear localization signal was identified in *c-rel* and *v-rel* by site-directed mutagenesis (Gilmore and Temin, 1986), and this positively charged region (positions 358–363) is maintained in NF-κB. A possible site for serine phosphorylation, Arg-Arg-X-Ser (at position 332 of NF-κB) (Steward, 1987), is maintained between all the proteins and might be involved in regulating conserved processes.

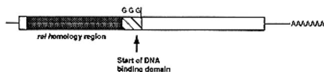
An important question that still remains to be answered is the identity and function of the carboxy-terminal half of

```
               AGCGGCCGCCGCGGGCGCGCTCTAGCAGCGCAGGCCGGAGCTCAGGGCCCCCGCGCGCCCGGCCCGCCCCCGCTTCTCCGCCCCGCGCCGCAGCCATGGCGCGCCGCTGAGC   110
        CGCCCCGCCCGCCCCGCCCGCGCCCCGACCCGGCTCGGCTCCCGCCGGTCCGCGCCGCTCCGCAGCGGAGCCCGCAGGCGAGGAGAGGCCGCGCCCATCTCCAGGGTACCCTCAGAGGCCAG   230
        AAGAGGGTGTCAGAGCCCTTGTAACTGGAGTTTGACGGTCGTGAGCTGCGCATCTTCACCATGGCAGACGATGATCCCTACGGAACTGGGCAAATGTTTCATTTGAACACTGCTTTGACT   350
                                                                                M  A  D  D  D  P  Y  G  T  G  Q  M  F  H  L  N  T  A  L  T
        CACTCAATATTTAATGCAGAATTATATTCACCAGAAATACCACTGTCAACAGATGGCCCATACCTTCAAATATTAGAGCAACCAAAACAGAGGGGATTTCGATTCCGCTATGTGTGTGAA   470
MOUSE   H  S  I  F  N  A  E  L  Y  S  P  E  I  P  L  S  T  D  G  P  Y  L  Q  I  L  E  Q  P  K  Q  R  G  F  R  F  R  Y  V  C  E
RABBIT                                                                                                             *  *  *  *  *
        GGCCCATCACACGGAGGGCTTCCGGGAGCCTCTAGTGAGAAGAACAAGAAATCCTACCCACAGGTCAAAATTTGCAACTATGTGGGGCCTGCAAAGGTTATCGTTCAGTTGGTCACAAAT   590
MOUSE   G  P  S  H  G  G  L  P  G  A  S  S  E  K  N  K  K  S  Y  P  Q  V  K  I  C  N  Y  V  G  P  A  K  V  I  V  Q  L  V  T  N
RABBIT  *  *  *  *  *  *  *  *  *  *  *  *  *  *                                         *  *  *  *  *  *  *  *  D
        GGAAAAAACATCCACCTGCACGCCCACAGCCTGGTGGGCAAGCACTGTGAGGACGGGGTATGCACCGTAACAGCAGGACCCAAGGACATGGTGGTTGGCTTTGCAAACCTGGGAATACTT   710
MOUSE   G  K  N  I  H  L  H  A  H  S  L  V  G  K  H  C  E  D  G  V  C  T  V  T  A  G  P  K  D  M  V  V  G  F  A  N  L  G  I  L
RABBIT  *  *  *  *  *  *  *  *                 *  *  *  *  *  *  I  *  *  *  N  *  *  *  *
        CATGTGACTAAGAAAAAGGTATTTGAAACACTGGAAGCACGGATGACAGAGGCGTGTATTAGGGGCTATAATCCTGGACTTCTGGTGCATTCTGACCTTGCCTATCTACAAGCAGAAGGC   830
MOUSE   H  V  T  K  K  K  V  F  E  T  L  E  A  R  M  T  E  A  C  I  R  G  Y  N  P  G  L  L  V  H  S  D  L  A  Y  L  Q  A  E  G
RABBIT                                                          *     *  *  *  *  *  *  *  *  *  P  *  *  *  *  *  *  *  T  *  *
        GGAGGAGACCGGCAACTCACAGACAGAGAGAAGGAGATCATCCGCCAGGCAGCCGTGCAGCAGACCAAGGAGATGGACCTGAGCGTGGTGCGCCTCATGTTCACAGCCTTCCTCCCTGAC   950
MOUSE   G  G  D  R  Q  L  T  D  R  E  K  E  I  I  R  Q  A  A  V  Q  Q  T  K  E  M  D  L  S  V  V  R  L  M  F  T  A  F  L  P  D
RABBIT  *  *  *
        AGCACTGGCAGCTTCACTCGGAGACTGGAGCCTGTGGTGTCAGACGCCATCTATGATAGCAAAGCCCCGAATGCATCCAACCTGAAAATCGTGAGAATGGACAGAACAGCAGGATGTGTG   1070
MOUSE   S  T  G  S  F  T  R  R  L  E  P  V  V  S  D  A  I  Y  D  S  K  A  P  N  A  S  N  L  K  I  V  R  M  D  R  T  A  G  C  V
RABBIT     *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *                                         *  *  *  *  *  *  *
        ACGGGAGGGGAGGAGATTTACCTTCTCTGTGACAAGGTTCAGAAAGATGACATCCAGATTCGGTTTTATGAAGAGGAAGAAAATGGCGGAGTTTGGGAAGGATTTGGGGACTTTTCCCCC   1190
MOUSE   T  G  G  E  E  I  Y  L  L  C  D  K  V  Q  K  D  D  I  Q  I  R  F  Y  E  E  E  E  N  G  G  V  W  E  G  F  G  D  F  S  P
RABBIT  *  *  *  *  *  *  *  *  *  *  *     *     *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  I  *  *  *  *  *  *  *  *  *
        ACGGATGTTCATAGACAGTTTGCCATTGTCTTCAAAACGCCAAAGTATAAGGATGTCAACATTACAAAGCCAGCTTCCGTGTTTGTTCAGCTTCGGAGGAAATCAGACCTGGAAACTAGT   1310
MOUSE   T  D  V  H  R  Q  F  A  I  V  F  K  T  P  K  Y  K  D  V  N  I  T  K  P  A  S  V  F  V  Q  L  R  R  K  S  D  L  E  T  S
RABBIT  *  *  *  *     *  *  *  *  *  *              *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
        GAACCGAAACCCTTTCTCTACTACCCTGAAATCAAAGACAAAGAGGAAGTGCAAAGGAAACGCCAGAAGCTTATGCCGAACTTCTCGGACAGCTTCGGCGGCGGCAGTGGAGCGGGAGCC   1430
MOUSE   E  P  K  P  F  L  Y  Y  P  E  I  K  D  K  E  E  V  Q  R  K  R  Q  K  L  M  P  N  F  S  D  S  F  G  G  G  S  G  A  G  A
RABBIT  *  *  *  *  *  *  *  *  *  *  *  *  *
        GGTGGTGGAGGCATGTTCGGTAGTGGCGGTGGCGGAGGGAGTACCGGAAGCCCTGGCCCAGGGTATGGCTACTCGAACTACGGATTTCCTCCCTACGGTGGGATTACATTCCATCCCGGA   1550
        G  G  G  G  M  F  G  S  G  G  G  G  G  G  S  T  G  S  P  G  P  G  Y  G  Y  S  N  Y  G  F  P  P  Y  G  G  I  T  F  H  P  G
        GTCACGAAATCCAACGCAGGGGTCACCCATGGCACCATAAACACCAAATTTAAAAATGGCCCTAAAGATTGTGCCAAGAGTGATGACGAGGAGAGTCTGACTCTCCCTGAGAAGGAAACT   1670
        V  T  K  S  N  A  G  V  T  H  G  T  I  N  T  K  F  K  N  G  P  K  D  C  A  K  S  D  D  E  E  S  L  T  L  P  E  K  E  T
        GAAGGTGAAGGGCCCAGCCTGCCCATGGCCTGCACCAAGACGGAACCCATCGCCTTGGCATCCACCATGGAAGACAAGGAGCAGGACATGGGATTTCAGGATAACCTCTTTCTCGAGAAG   1790
        E  G  E  G  P  S  L  P  M  A  C  T  K  T  E  P  I  A  L  A  S  T  M  E  D  K  E  Q  D  M  G  F  Q  D  N  L  F  L  E  K
        GCTCTGCAGCTCGCCAGGCGACACGCCAACGCCCTTTTCGACTACGCAGTGACGGGGGATGTGAAGATGTTGCTGGCCGTGCAACGCCATCTCACCGCCGTGCAGGATGAGAATGGGGAC   1910
        A  L  Q  L  A  R  R  H  A  N  A  L  F  D  Y  A  V  T  G  D  V  K  M  L  L  A  V  Q  R  H  L  T  A  V  Q  D  E  N  G  D
        AGTGTCTTACACTTAGCCATCATCCACCTCCACGCTCAGCTCGTGAGGGATCTGCTGGAAGTCACATCTGGTTTGATCTCTGATGACATCATCAACATGAGAAATGACCTGTATCAGACA   2030
        S  V  L  H  L  A  I  I  H  L  H  A  Q  L  V  R  D  L  L  E  V  T  S  G  L  I  S  D  D  I  I  N  M  R  N  D  L  Y  Q  T
        CCTCTGCACTTGGCCGTGATCACCAAGCAGGAAGATGTAGTAGAGGATTTGCTGAGGGTTGGGGCTGACCTGAGCCTTCTGGACCGCTGGGGCAACTCTGTCCTGCACCTAGCTGCCAAA   2150
        P  L  H  L  A  V  I  T  K  Q  E  D  V  V  E  D  L  L  R  V  G  A  D  L  S  L  L  D  R  W  G  N  S  V  L  H  L  A  A  K
        GAAGGACACGACAGAATCCTCAGCATCCTGCTCAAGAGCAGAAAAGCAGCGCCCCTTATCGACCACCCCAATGGGGAAGGTCTAAATGCCATCCACATAGCTGTGATGAGCAATAGCCTG   2270
        E  G  H  D  R  I  L  S  I  L  L  K  S  R  K  A  A  P  L  I  D  H  P  N  G  E  G  L  N  A  I  H  I  A  V  M  S  N  S  L
        CCATGTCTGCTGCTGCTGGTGGCTGCCGGGGCAGAAGTCAATGCTCAGGAGCAGAAGTCTGGGCGCACGCCGCTGCACCTGGCCGTGGAGTACGACAACATCTCCTTGGCTGGCTGCCTG   2390
        P  C  L  L  L  L  V  A  A  G  A  E  V  N  A  Q  E  Q  K  S  G  R  T  P  L  H  L  A  V  E  Y  D  N  I  S  L  A  G  C  L
        CTTCTGGAGGGTGATGCCCACGTGGACAGTACCACCTATGATGGGACTACACCTCTGCATATAGCGGCCGGAAGAGGGTCCACCAGACTGGCAGCTCTTCTCAAAGCAGCAGGAGCAGAC   2510
        L  L  E  G  D  A  H  V  D  S  T  T  Y  D  G  T  T  P  L  H  I  A  A  G  R  G  S  T  R  L  A  A  L  L  K  A  A  G  A  D
        CCCCTGGTGGAGAACTTTGAGCCTCTCTATGACCTGGACGACTCTTGGGAGAAGGCTGGAGAAGATGAGGGAGTGGTGCCAGGTACCACACCCCTGGACATGGCTGCCAACTGGCAGGTA   2630
        P  L  V  E  N  F  E  P  L  Y  D  L  D  D  S  W  E  K  A  G  E  D  E  G  V  V  P  G  T  T  P  L  D  M  A  A  N  W  Q  V
        TTTGACATACTAAATGGGAAACCGTATGAGCCTGTGTTCACATCTGATGATATACTACCACAAGGGGACATGAAGCAGCTGACAGAAGACACGAGGCTACAACTCTGCAAACTGCTGGAA   2750
        F  D  I  L  N  G  K  P  Y  E  P  V  F  T  S  D  D  I  L  P  Q  G  D  M  K  Q  L  T  E  D  T  R  L  Q  L  C  K  L  L  E
        ATTCCTGATCCAGACAAAAACTGGGCCACTCTGGCACAGAAGTTGGGTCTGGGGATATTGAACAATGCCTTCCGGCTGAGTCCTGCTCCTTCTAAAACTCTCATGGACAACTATGAGGTC   2870
        I  P  D  P  D  K  N  W  A  T  L  A  Q  K  L  G  L  G  I  L  N  N  A  F  R  L  S  P  A  P  S  K  T  L  M  D  N  Y  E  V
        TCTGGGGGTACCATCAAAGAGCTGATGGAGGCCCTGCAACAGATGGGCTACACAGAGGCCATTGAAGTGATCCAGGCAGCCTTCCGCACCCCGGCAACCACAGCCTCCAGCCCCGTGACC   2990
        S  G  G  T  I  K  E  L  M  E  A  L  Q  Q  M  G  Y  T  E  A  I  E  V  I  Q  A  A  F  R  T  P  A  T  T  A  S  S  P  V  T
        ACTGCTCAGGTCCACTGTCTGCCTCTCTCGTCTTCCTCCACGAGGCAGCACATAGATGAACTCCGGGATAGTGACAGCGTCTGTGACAGTGGTGTGGAGACATCCTTCCGCAAACTCAGC   3110
        T  A  Q  V  H  C  L  P  L  S  S  S  S  T  R  Q  H  I  D  E  L  R  D  S  D  S  V  C  D  S  G  V  E  T  S  F  R  K  L  S
        TTTACAGAGTCTCTTACTGGAGACAGCCCACTGCTATCTCTGAACAAAATGCCCCACGGTTATGGGCAGGAAGGACCTATTGAAGGCAAAATTTAGCCTGCTGGCCGTTCCCCCACACTG   3230
        F  T  E  S  L  T  G  D  S  P  L  L  S  L  N  K  M  P  H  G  Y  G  Q  E  G  P  I  E  G  K  I  •
        TGTAAACCAAAGCCCTGACAGTCCATTGCATCGTCCCAAAGGAGGAAGGCAAACCGAATCCAAAGGTGCTGGAGAATCGCCGGCCTGCAGGGTCACTCGATTTCATTCAAGGCCTTCCGA   3350
        ATTTGGCGTCCTTCTTGGTTCTGAAATGAAATGTAGTTGCCACGCACAGACGGTGTCTAGCAATCATGGCGCTCCCTCGCTCAGCTGCACTCTATGGCTCAGGTGCAGTGTCTTGAGCTT   3470
        TCTCTGCTGCTACTGGATCACATTTGCTTTGTGTTGTTACTGCTGTCCCTCCGCTGGGTTCCTGCTGTCATTAAAAGGTGTCGCTGTCCCCACCCGGTGTCCTTTCTAGCCATCTACTGT   3590
        AAGTTGTGCATTCAAATTAAGATTAAGGAAAAACATATTTTTAAATGAGTACCTTGATGCGCAATAAAAAAAAAGACATTTCTTTTTTTAATGTGGTTTATCTGTGATTTAAAAATAAAA   3710
        AACACATGAACTTATCAATATTTAAAACATGCTACAATCAGTGNTGAAAATAGTATTTTCCCCGTTTTATGCATTTTACATTTGTAAATATGTTTTCTAATCAATACTTTAAAAGAAGAA   3830
        TGTTGAATTTATAAAATGCTATTTACTTTTTTATTTAT AATAAAGTACAGCACATGTGACTAAAAA...3896
```

Figure 2. Nucleotide and Amino Acid Sequence of the DNA Binding Subunit of NF-κB

The nucleotide sequence of clone 3, with the deduced amino acid sequence (single-letter code) of the major open reading frame, is shown. The sequences obtained from rabbit NF-κB tryptic peptides are shown below the sequence; asterisks represent identical residues, and changes between the rabbit and mouse sequences are indicated. The putative polyadenylation sequence, AATAAA, is underlined. Only the first 5 bases of the poly(A) tail are shown.

A.



B.

Figure 3. Structure and Homology of the NF-κB Clone

(A) Schematic representation of the NF-κB clone. The long open reading frame of 971 amino acids is indicated by the box. The shaded region represents the region of homology with *rel* and *dorsal*. The approximate end for a 50 kd protein is indicated with the dashed line. The glycine-rich region (GGG) is indicated.

(B) Comparision of the NF-κB p50 sequence with that of *dorsal*, v-*rel*, and turkey, mouse, and human c-*rel*. Only the region of homology is indicated. Identical residues are in open boxes; conservative changes are indicated by stippling. The sequence of mouse c-*rel* is from Bull et al. (1990).

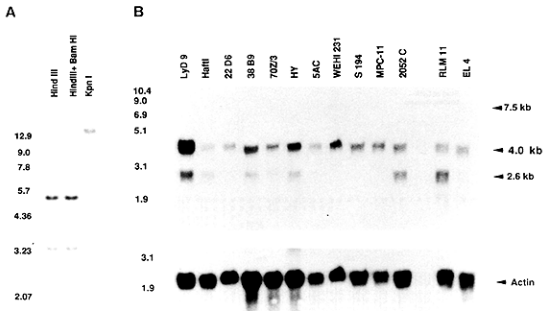


Figure 4. DNA and RNA Analysis with the NF-κB clone

(A) Southern blot analysis of mouse genomic DNA. Genomic DNA prepared from mouse cells was digested with the indicated restriction enzymes, electrophoresed through a 0.8% agarose gel, and transferred to Zetabind membrane. Hybridization was carried out at 42°C with 50% formamide (see Experimental Procedures). The blot was probed with a labeled fragment corresponding to positions 629 to 1348.

(B) Expression of the p50 mRNA in different cells. Sixteen micrograms of once-selected poly(A)$^+$ RNA or 8 μg of twice-selected poly(A)$^+$ RNA (22D6) was isolated from the indicated cell lines by standard techniques and electrophoresed through a formaldehyde–agarose gel and then transferred to Zetabind. The blot was probed with the labeled 4.0 kb cDNA.

the protein. Because we cannot detect the p65 peptide sequences in the carboxy-terminal half, we believe that p65 is the product of a separate gene. The size of c-*rel*, which is 68 kd (Grumont and Gerondakis, 1989; Brownell et al. 1989; Simek and Rice, 1988b), is provocatively similar to that of the p65 protein (65 kd); therefore we specifically compared the p65 protein sequences with the mouse and human c-*rel* sequences. The comparision indicates that p65 is probably another member of the *rel* family of proteins (the sequence of peptide 1 from p65 is identical to c-*rel* between positions 238 and 246) but is probably not c-*rel* itself (based on sequences from the other p65 peptides). Cloning of this protein will help to clarify its relation to the other *rel* proteins.

We have also compared the sequence of p50 NF-κB from mouse with the sequence of the human p50 NF-κB/KBF-1 (see Kieran et al., 1990). The comparision indicates over 95% homology between the two proteins except for a stretch of about 60 amino acids between positions 432 and 490. Interestingly, this region of difference is also a segment that divides the open reading frame into an amino-terminal half coding for p50 and a carboxy-terminal polypeptide of about 53 kd, and the intervening sequence could therefore be a region for proteolytic processing. The degree of conservation in the carboxy-terminal region suggests that it could encode a functional protein; future experiments with the recombinant proteins should answer this question.

Southern and Northern Analysis of the Gene for p50

Southern blot analysis of mouse genomic DNA digested with KpnI, using a portion of the cloned cDNA as probe, revealed a single band. Hybridization of DNA digested with HindIII produced two hybridizing fragments, of 5.5 kb and 3.3 kb. The pattern was not changed if digestion was carried out with BamHI in addition to HindIII (Figure 4A). These results indicate that either a single gene or, less likely, very closely linked genes are detected under the stringent conditions of hybridization used.

We also examined the expression of p50 mRNA in various cell lines, including a number of lymphoid cell lines from different developmental stages (Figure 4B). All of these cell lines expressed two mRNA species, of 4.0 kb and 2.6 kb, in widely varying amounts. It is clear that we have cloned the 4.0 kb species, and our failure to obtain clones for the 2.6 kb mRNA is probably a reflection of its far lower abundance in mouse 22D6 cells, from which the cDNA library was made. Although we cannot rule out the possibility completely, the virtual absence of the smaller RNA in 22D6, WEHI231, and S194 cell lines, all of which contain readily detectable NF-κB protein, argues against the 2.6 kb mRNA coding for NF-κB. We are, however, trying to clone this mRNA to determine its actual role and its relation to NF-κB. The 4.0 kb mRNA is expressed in all cell types tested; interestingly, the pro-B cell line LyD9 appears to express higher levels of the p50 mRNA than do

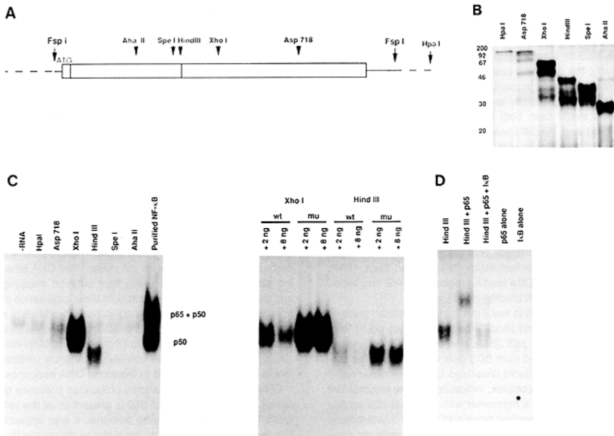


Figure 5. Translation and DNA Binding of the NF-κB Clone

(A) Schematic representation of the NF-κB clone linearized with the indicated restriction enzymes. HpaI linearizes the clone by cutting the vector, pSP73. The linearized templates were used for transcription employing SP6 RNA polymerase.

(B) In vitro translated products from the FspI fragment linearized with different restriction enzymes. Approximately 4% of each translation reaction (50 μl) was analyzed on a 15% SDS–polyacrylamide gel. Following electrophoresis, the gel was fixed, treated with Autofluor (National Diagnostics), dried, and exposed for fluorography.

(C) Binding of the in vitro translated proteins to the κB site. Seven microliter samples of the HpaI and Asp718 translation reactions, and 3 μl samples of the remaining translation reactions, were used for determining binding to the labeled κB probe using electophoretic mobility shift analysis. The right side shows the specificity of complex formation by the in vitro translated proteins. Competition analysis was done using 2 ng and 8 ng of oligonucleotides containing either a wild-type (wt) or mutant (mu) κB site.

(D) Heterodimer formation with p65 and inhibition by IκB. Three microliters of the HindIII translation reaction was used alone or with gel-purified and renatured p65. Highly purified IκB was also added to a reaction mix containing the recombinant protein and p65 and incubated as described (Baeuerle and Baltimore, 1988b). The proteins were then analyzed for binding to the κB probe using electrophoretic mobility shift analysis (Ghosh and Baltimore, 1990).

the other pre-B and mature B cells. Also, the 7.5 kb mRNA for c-*rel* can be detected as a faint band in mature B cells and plasmocytomas (not visible at the exposure in Figure 4B) (Brownell et al., 1987).

Functional Identity of the p50 NF-κB Clone

Because of the discrepancy between the size of the purified p50 NF-κB polypeptide and the predicted size of the protein encoded by this clone, we wanted to verify the size of the protein made in an in vitro transcription–translation experiment. The cDNA, cloned into a plasmid vector, was used as a template for transcription driven by T7 RNA polymerase. The transcripts from the native clone translated very poorly but did result in a protein of around M_r 107,000 as well as a number of smaller species. The 5′ untranslated region of the clone was extremely rich in G and C residues, a feature that has been suggested to result in inefficient translation. Therefore, we subcloned a FspI fragment that removed most of the untranslated sequences and used it for our subsequent in vitro transcription–translation experiments. We also generated smaller versions of the FspI clone by linearizing the plasmid with the restriction enzymes Asp718, XhoI, HindIII, SpeI, and AhaII (Figure 5A). Translation of the mRNAs produced from these templates gave proteins of the expected sizes (Figure 5B), although the intact cDNA (plasmid cut with HpaI) and the Asp718-linearized clones translated poorly.

The translated proteins were then used in an electrophoretic mobility shift assay with the κB probe, to determine if they could bind to DNA. The protein produced from

the intact clone (linearized with HpaI) did not appear to bind to DNA, while the protein from the Asp718 fragment (which is shorter by 205 amino acids, giving a protein of M_r 85,000) bound poorly to DNA (Figure 5C). (Alternatively, the observed binding may be due to some of the smaller proteins present, which are probably prematurely terminated products; see Figure 5B.) However, the proteins made from both the XhoI-linearized (498 amino acids) and HindIII-linearized (363 amino acids) clones bound specifically with high affinity to the κB site, but shortened proteins made from SpeI-linearized (338 amino acids) or AhaII-linearized (255 amino acids) clones did not bind DNA, indicating that the beginning of the DNA binding motif is localized between the HindIII and the SpeI sites (Figures 3A and 5C). However, we could not discern any apparent homology to any known DNA binding motif such as the zinc finger, leucine zipper, helix-turn-helix, helix-loop-helix, or homeobox (Mitchell and Tjian, 1989). Accordingly, the DNA binding region of NF-κB may represent a novel DNA binding motif.

We then wished to see if the in vitro translated proteins from the XhoI- and HindIII-linearized clones could form heteromers with p65 (Baeuerle and Baltimore, 1989). When p65 purified from SDS gels was added to the proteins from the HindIII-linearized clone, we observed a slower-migrating complex, indicating that the recombinant protein can form a heteromer with pure p65 (the protein from the XhoI-linearized clone also forms the heteromeric complex, which is partly masked by the homodimeric complex; data not shown) (Figure 5D). In addition, we find that IκB can block the DNA binding activity of only the recombinant heteromeric complex, in amounts similar to that required for the native proteins (Figure 5D) (Baeuerle and Baltimore, 1989). These results indicate that the cloned p50 NF-κB displays properties identical to the purified factor.

Discussion

We have isolated and sequenced a mouse cDNA clone for the DNA binding subunit of the transcription factor NF-κB. Surprisingly, the cDNA predicts an open reading frame of 971 amino acids giving a protein of M_r 107,000, significantly larger than the purified DNA binding subunit (M_r 50,000). Nearly 180 amino acids of protein sequence obtained by sequencing tryptic peptides from p50 are found at the amino-terminal half of the open reading frame, which should therefore encode the DNA binding subunit. This suggests a novel maturation process for an intracellular cytosolic protein, where it is initially synthesized as a larger protein but then gets trimmed to its functional size. It is difficult to predict if this process has a regulatory role or if the carboxy-terminal half encodes a functional polypeptide. Protein sequence from p65, the other subunit of NF-κB, is not evident in the carboxy-terminal half of the p50 NF-κB clone, but it could still code for IκB, especially because the presence of the carboxyl terminus of the protein appears to inhibit DNA binding. Distinguishing between these possibilities will require antibodies raised against different regions of the open reading frame.

The full-length protein of 107,000 daltons does not bind to the κB site. This could be due to a segment either interfering with the homodimerization or acting to inhibit DNA binding. When about 500 carboxy-terminal amino acids are removed from the protein, it then binds to DNA with high affinity, thus localizing the interfering domain to the carboxyl terminus. It is quite possible that the full-length form of p50 NF-κB is present in many cells, but cannot be detected because the only assay for NF-κB has been one based on DNA binding (Sen and Baltimore, 1988a). Because deletion of the cDNA clone beyond amino acid 360 prevents DNA binding, it appears that the region from 363 to 338 contains the start of the DNA binding region. However, we cannot detect any known DNA binding motif in this region, including the leucine zipper, helix-turn-helix, homeobox, helix-loop-helix, or zinc finger, or even a highly basic region (Mitchell and Tjian, 1989). Therefore, NF-κB probably uses a novel and as yet unidentified DNA binding motif. Further studies using finer deletion mapping and site-directed mutagenesis should allow localization of this domain. It is also interesting to note that MBP-1, a DNA binding protein that binds specifically to the κB site, uses zinc fingers for binding to DNA (Fan and Maniatis, 1989). Thus, it appears that completely different binding motifs can be used to bind to the same DNA sequence.

Consistent with the apparently ubiquitous presence of NF-κB, the 4.0 kb mRNA of p50 is present in all the cell types tested, albeit in varying amounts. It also appears that mature B cells do not contain significantly greater levels of transcripts compared with earlier developmental stages, implying that the change of NF-κB from an inducible to a constitutive form is not accompanied by an up-regulation of transcription. Beacuse of the high degree of homology between p50 NF-κB and other *rel* proteins, it was likely that we would detect these related mRNAs on a Northern blot. Earlier studies using mouse c-*rel* as a probe detected the 7.5 kb mRNA for c-*rel* but also cross-hybridized to a 4.0 kb mRNA that was present in spleen, muscle, and hematopoietic tissues and to a 2.6 kb mRNA whose distribution was more uneven (Brownell et al., 1987). It now appears likely that the 4.0 kb mRNA is that encoding p50 NF-κB, while the 2.6 kb mRNA probably encodes a tissue-specific *rel*-homologous protein. The unique arrangement of the p50 NF-κB mRNA, where the 3′ region apparently encodes an inhibitory domain, also explains the failure to obtain NF-κB clones using oligonucleotide screening (Singh et al., 1988; Fan and Maniatis, 1989). Oligo(dT)-primed cDNA libraries would not work because they would include the carboxy-terminal inhibitory domain.

The most remarkable feature of the p50 NF-κB clone is its homology over an extensive region with v-*rel* (Stephens et al., 1983), its cellular counterpart c-*rel* (Wilhelmsen et al., 1984; Grumont and Gerondakis, 1989; Brownell et al., 1989), and Drosophila *dorsal* (Steward, 1987). v-*rel* is the oncogene carried by the reticuloendotheliosis virus strain T, an acute leukemia virus in turkey. The v-*rel* protein is an *env*–*rel*–*env* fusion, containing 11 *env*-derived amino acids, 474 c-*rel* amino acids, and 18 amino acids at the carboxyl terminus that are derived from out-of-frame translation of

env (Stephens et al., 1983; Wilhelmsen et al., 1984). The v-*rel* protein is not homologous to any other oncogene. Multiple nucleotide differences between v-*rel* and turkey c-*rel* result in 14 amino acid changes and some small deletions. These differences have been shown to be important for transformation and immortalization (Sylla and Temin, 1986). v-*rel* is present primarily as a soluble cytosolic protein in spleen cells, but is largely a nuclear protein in nontransformed fibroblasts (Gilmore and Temin, 1986). However, subcellular localization appears to be immaterial to transforming capability (Gilmore and Temin, 1988). The p59$^{v\text{-}rel}$ protein is normally found associated with a number of polypeptides: one is a serine/threonine protein kinase (Walro et al. 1987), while another has been identified as a heat shock protein (Lim et al., 1990). The functions of the other proteins in the complex are unknown.

Even less is known about the function of c-*rel*, which has been cloned from turkey, mouse, and human cells. The product of the c-*rel* gene is a 68,000 dalton protein that is present in the cytosol and is probably complexed with proteins similar to those bound to v-*rel* (Morrison et al., 1989; Simek and Rice, 1988a). The associated proteins might act as anchors for keeping c-*rel* localized to the cytosol. The carboxy-terminal portion of c-*rel* has potent *trans*-activation potential (Hannink and Temin, 1989; Kamens et al., 1990; Bull et al., 1990), suggesting that it probably functions in the nucleus and is inactive when present in the cytosol.

The extent of similarity in the *rel*-homologous region between *rel* and p50 NF-κB suggests that the two proteins probably have similar structures in this region and perform some common function. Because the carboxy-terminal halves of these proteins diverge almost completely, it is likely that specific functions for these proteins are determined by these divergent regions. For c-*rel* no DNA binding has been found after significant effort, so that it may be involved only in *trans*-activation (Hannink and Temin, 1989; Bull et al., 1990). Shared functions, such as nuclear–cytosolic translocation, are probably specified in the common *rel*-homologous region.

The *dorsal* gene product is probably a direct regulator of dorsal–ventral polarity (Steward and Nüsslein-Volhard, 1986). It is one of a number of maternal effect genes acting to establish the polarity of the embryo. Selective nuclear transport of *dorsal* may be basis of the ability of this morphogen to establish a gradient of response (Roth et al., 1989; Rushlow et al., 1989; Steward, 1989). It has been suggested that selective nuclear transport is determined by a postranslational event such as proteolysis (Rushlow et al., 1989). This is provocative because two genes upstream of *dorsal*—*easter* and *snake*—probably encode serine proteases (Chasan and Anderson, 1989; DeLotto and Spierer, 1986). If a proteolytic event does occur, it might involve the removal of an inhibiting carboxy-terminal domain similar to that in the clone for p50 NF-κB. Such a domain could mask a nuclear translocation signal and be equivalent to IκB. The analogy between *dorsal* and NF-κB can be extended further if the gene product of *cactus* acts as an anchor for the *dorsal* protein in the cytoplasm (Roth et al., 1989). It is known that in the absence of *cactus* function, *dorsal* becomes exclusively localized to the nucleus; therefore the *cactus* gene product could inhibit nuclear localization by *dorsal*–*cactus* interaction.

The role of p65 in the NF-κB complex remains obscure. The protein sequence for p65 indicates that it is related to the c-*rel* protein (Grumont and Gerondakis, 1989). Because the nonhomologous carboxyl terminus of c-*rel* can function as a transcriptional activator, it is probable that p65 can also perform a similar function (Bull et al., 1990). Thus, p65 and p50 could associate with one another via the *rel*-homology region, and while p50 would bind to DNA, p65 would act as a transcriptional activator. The similarity between p65 and c-*rel* also suggests that p50 could form such heteromeric interactions with other *rel* proteins, which would act as *trans*-activators for different genes. For instance, HIVEN 86 is a protein that associates with NF-κB (Böhnlein et al., 1988), and according to this model could be a member of the *rel trans*-activators. Therefore the *rel* region would act as a dimerization motif, similar to the role played by the helix-loop-helix motif in allowing different proteins such as MyoD, Drosophila *achaete-scute* and *daughterless*, and the immunoglobulin enhancer E box proteins, E12 and E47, to form heterodimers with one another (Murre et al., 1989). Cloning of p65 and further experiments using specific antibodies should help to answer some of these questions.

Experimental Procedures

Purification of NF-κB

NF-κB was purified from rabbit lungs essentially as described before (Ghosh and Baltimore, 1990), with the following modifications. The NF-κB–IκB complex was purified up to the Sephacryl S-200 HR stage, after which it was dissociated with sodium deoxycholate and the individual components were separated on a Q-Sepharose anion exchange column. The separated NF-κB was then purified by chromatography on an S-Sepharose column. The eluted protein was then adjusted to 0.1 M KCl and mixed with calf thymus DNA at 10 μg/ml for 30 min before loading on a column containing bound oligonucleotide with the κB site (GGGGACTTTCC). The column was washed with the equilibrating buffer followed by 0.2 M KCl buffer, and the bound NF-κB was eluted with buffer containing 0.4 M KCl. The 0.4 M eluate was adjusted to 0.1 M KCl and loaded on a column containing oligonucleotides with a mutant κB site (GGCGACTTTCC), and all the NF-κB protein now eluted at 0.2 M KCl. The eluted protein was rechromatographed on another affinity column containing wild-type κB sites, and the final 0.4 M KCl eluate was used for subsequent experiments.

Protein Sequence Analysis following Tryptic Digestion

Purified protein (20 μg) was concentrated by precipitation with trichloroacetic acid and resuspended in SDS sample buffer. It was then loaded on a 1.5 mm thick, 8.75% SDS–polyacrylamide gel in a wide lane. Following electrophoresis, the proteins on the gel were transferred onto a nitrocellulose membrane in a Hoefer blotting apparatus at 30 V for 12 hr. The transfer buffer contained 25 mM Tris, 192 mM glycine, and 20% methanol. Proteins were visualized on the membrane by Ponceau S staining, and the predominant 50 and 65 kd bands were excised. In situ digestion of the proteins with 2 μg of trypsin at 37°C for 15 hr was performed exactly as described previously (Tempst et al., 1990; Aebersold et al., 1987). An enzyme blank was done on a strip of nitrocellulose cut from a blank area of the same blot. The tryptic peptides released into the supernatant were reduced by 0.2% β-mercaptoethanol for 30 min at 37°C and S-alkylated using 0.6% 4-vinyl pyridine (10% solution in ethanol). S-alkylation was done for 30 min at room temperature under argon and in the dark (Tempst et al., 1990). The mixture was then injected for analysis on a narrow-bore reverse-phase HPLC system consisting of a microgradient pump (Brownlee

Labs) and a 1000S diode array detector (Applied Biosystems). Identification of peptides containing Trp, Tyr, and pyridylethyl-cysteine was done by absorbance ratio analysis at 253, 277, and 297 nm (Casteels et al., 1990; Tempst et al., 1990). A 2.1 × 20 mm Aquapore RP-300 column (Applied Biosystems) was used for primary separations; repurification of peak fractions was done on a 2.1 × 250 mm Vydac 218TP52 C18 column. Gradient I (RP-300 column) was 60 min isocratic at 5% B, linear 5%–50% B in 45 min, linear 50%–100% B in 20 min; gradient II (C18 column) was the same except that the isocratic step was 15 min. Solvent A was 0.1% trifluoroacetic acid; solvent B was 0.09% trifluoroacetic acid in 70% acetonitrile. Fractions were collected manually and stored at −20°C.

Peptides were sequenced with the aid of an Applied Biosystems model 477A automated sequenator, operated according to Hewick et al. (1981). Stepwise liberated PTH-amino acids were identified using an on-line 120A HPLC system equipped with a PTH C18 column (Applied Biosystems). The standard Applied Biosystems method was optimized for subpicomole PTH analysis as described (Tempst and Riviere, 1989).

Isolation of Recombinant Clones Encoding the p50 Subunit of NF-κB

Oligonucleotide primers deduced from the first 7 amino acids of peptides T7.1 and T24 were synthesized and used as primers. Total RNA was isolated from fresh rabbit lung and spleen using guanidinium thiocyanate as denaturant. One microgram of the RNA was used for each PCR reaction. The primers were used in similar amounts, although the degeneracy of the primer from peptide T24 was greater. The reactions were carried out under conditions as described (Schlissel and Baltimore, 1989), with the first five cycles having an annealing temperature of 48°C and the subsequent 35 cycles at 60°C. One-tenth of the reactions were analyzed, and the amplified fragment detected with primers 2 and 3 was isolated and reamplified. The reamplified fragment was cut with EcoRI and HindIII and directionally cloned into pBSK⁺ and pBSK⁻ plasmids. The cloned fragment was then sequenced from both ends on single-stranded templates using the Sanger dideoxy method (Sanger et al., 1977) with Sequenase.

The cloned fragment was excised from the plasmid and was radiolabeled using the random hexamer priming method (Feinberg and Vogelstein, 1983). The screening of the mouse 22D6 cDNA library (kindly provided by M. Oettinger, C. Gorka, and D. Schatz) was done essentially as described by Schatz et al. (1989), except that the final washes were done using 1× SSC, 0.1% SDS at 68°C. Five independent clones were isolated and cloned into the NotI site of pBSK⁺ and pBKS⁻. The inserts were characterized by restriction enzyme mapping, and one of the clones, clone 3, was used for further study.

Clone 3 was sequenced on both strands using subclones generated by exonuclease III digestion. Occasional "chewing back" by exonuclease III removed the sequences complementary to the sequencing primers in about 20% of the clones. This resulted in gaps in the sequence, which were subsequently covered by using specific sequencing primers (20-mer oligonucleotides) based on sequences already obtained from the exonuclease III deletion clones.

Southern and Northern Analysis

Southern blot analysis on mouse genomic DNA cut with EcoRI, HindIII, KpnI, and BamHI plus HindIII was done essentially as described (Schatz et al., 1989). The Northern blot was made with poly(A)⁺-containing RNA made from the different cell lines as indicated (Oettinger et al., 1990). Following hybridization with the probe (the 4.0 kb cDNA), the blot was washed under increasingly stringent conditions until the final wash with 0.4× SSC, 0.1% SDS at 68°C.

In Vitro Transcription and Translation

RNA was synthesized from the linearized plasmids by using either T7 (for Bluescript vectors) or SP6 (for the pSP73 vector) with the cap analog to produce capped transcripts. The RNA was extracted with phenol–chloroform and ethanol precipitated. The total RNA was suspended in 20 μl of water, and 2 μl of it was used to direct translation in a micrococcal nuclease–treated wheat germ extract (Promega). Proteins were synthesized at 25°C for 1 hr.

Acknowledgments

We would like to thank Dr. Alain Israël for sharing the sequence for the human p50 NF-κB/KBF-1 with us before publication, and the reviewers for their careful analysis of our manuscript. We would also like to thank Marjorie Oettinger, Dr. David G. Schatz, and Dr. Mark Schlissel for their material and intellectual help; Drs. X.-H. Sun, D. Black, and M. Feinberg for helpful advise and discussions; and M. Paskind and Dr. R. A. VanEtten for help with photography. This work was supported by Public Health Service grant GM39458 (to D. B.). S. G. was supported by a postdoctoral fellowship from the Irvington Institute for Medical Research, and G. P. N. was supported by a fellowship from the National Institutes of Health.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 10, 1990; revised July 30, 1990.

References

Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. H. (1987). Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after *in situ* protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA *84*, 6970–6974.

Baeuerle, P. A., and Baltimore, D. (1988a). Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor. Cell *53*, 211–217.

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

Baeuerle, P. A., and Baltimore, D. (1989). A 65 kD subunit of active NF-κB is required for inhibition of NF-κB by IκB. Genes Dev. *3*, 1689–1698.

Baeuerle, P. A., and Baltimore, D. (1990). The physiology of the NF-κB transcription factor. In Molecular Aspects of Cellular Regulation, Vol. 6, Hormonal Regulation of Transcription, P. Cohen and J. G. Foulkes, eds. (Amsterdam: Elsevier-Biomedical), in press.

Böhnlein, E., Lowenthal, J. W., Siekevitz, M., Ballard, D. W., Franza, B. R., and Greene, W. C. (1988). The same inducible nuclear protein(s) regulates mitogen activation of both the interleukin-2 receptor-alpha gene and type 1 HIV. Cell *53*, 827–836.

Brownell, E., Mathieson, B., Young, H. A., Keller, J., Ihle, J. N., and Rice, N. R. (1987). Detection of c-*rel* related transcripts in mouse hematopoietic tissues, fractionated lymphocyte populations, and cell lines. Mol. Cell. Biol. *7*, 1304–1309.

Brownell, E., Mittereder, N., and Rice, N. R. (1989). A human *rel* proto-oncogene cDNA containing an Alu fragment as a potential coding exon. Oncogene *4*, 935–942.

Bull, P., Morley, K. L., Hoekstra, M. F., Hunter, T., and Verma, I. M. (1990). The mouse c-rel protein has an N-terminal regulatory domain and a C-terminal transcriptional transactivation domain. Mol. Cell. Biol. *10*, in press.

Casteels, P., Ampe, C., Riviere, L., Van Damme, J., Elicone, C., Fleming, M., Jacobs, F., and Tempst, P. (1990). Isolation and characterization of abaecin, a major antibacterial response peptide in the honeybee. Eur. J. Biochem. *187*, 381–386.

Chasan, R., and Anderson, K. V. (1989). The role of *easter*, an apparent serine protease, in organizing the dorsal–ventral pattern of the Drosophila embryo. Cell *56*, 391–400.

DeLotto, R., and Spierer, P. (1986). A gene required for the specification of dorsal–ventral pattern in *Drosophila* appears to encode a serine protease. Nature *323*, 688–692.

Dynan, W. S. (1989). Modularity in promoters and enhancers. Cell *58*, 1–4.

Fan, C.-M., and Maniatis, T. (1989). A DNA binding protein containing

two widely separated zinc finger motifs that recognize the same DNA sequence. Genes Dev. *4*, 29–42.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Ghosh, S., and Baltimore, D. (1990). Activation *in vitro* of NF-κB by phosphorylation of its inhibitor IκB. Nature *344*, 678–682.

Gilmore, T. D., and Temin, H. M. (1986). Different localization of the product of the v-*rel* oncogene in chicken fibroblasts and spleen cells correlates with transformation by REV-T. Cell *44*, 791–800.

Gilmore, T. D., and Temin, H. M. (1988). *V-rel* oncoproteins in the nucleus and in the cytoplasm transform chicken spleen cells. J. Virol. *62*, 703–714.

Grumont, R. J., and Gerondakis, S. (1989). Structure of a mammalian c-*rel* protein deduced from the nucleotide sequence of murine cDNA clones. Oncogene Res. *4*, 1–8.

Hannink, M., and Temin, H. M. (1989). Transactivation of gene expression by nuclear and cytoplasmic rel proteins. Mol. Cell. Biol. *9*, 4323–4336.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981). A gas liquid solid phase peptide and protein sequenator. J. Biol. Chem. *256*, 7990–7997.

Kamens, J., Richardson, P., Mosialos, G., Brent, R., and Gilmore, T. (1990). Oncogenic transformation by vRel requires an amino-terminal activation domain. Mol. Cell. Biol. *10*, 2840–2847.

Kawakami, K., Scheidereit, C., and Roeder, R. G. (1988). Identification and purification of a human immunoglobulin enhancer binding protein (NF-κB) that activates transcription from a human immunodeficiency virus type 1 promoter *in vitro*. Proc. Natl. Acad. Sci. USA *85*, 4700–4704.

Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M. B., Kourilsky, P., Baeuerle, P. A., and Israël, A. (1990). The DNA binding subunit of NF-κB is identical to factor KBF1 and homologous to the *rel* oncogene product. Cell *62*, this issue.

Kozak, M. (1989). The scanning model for translation: an update. J. Cell Biol. *108*, 229–241.

Larson, J. S., Schuetz, T. J., and Kingston, R. E. (1988). Activation in vitro of sequence-specific DNA binding by a human regulatory factor. Nature *335*, 372–375.

Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M., and Caskey, C. T. (1988). Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. Science *239*, 1288–1291.

Lenardo, M. J., and Baltimore, D. (1989). NF-κB: a pleiotropic mediator of inducible and tissue-specific gene control. Cell *58*, 227–229.

Lim, M. Y., Davis, N., Zhang, J., and Bose, H. R. (1990). The v-*rel* oncogene product is complexed with cellular proteins including its proto-oncogene product and heat shock protein 70. Virology *175*, 149–160.

Mitchell, P. J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Morrison, L. E., Kabrun, N., Mudri, S., Hayman, M. J., and Enrietto, P. J. (1989). Viral rel and cellular rel associate with cellular proteins in transformed and normal cells. Oncogene *4*, 677–683.

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H., and Baltimore, D. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell *58*, 537–544.

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517–1523.

Picard, D., and Yamamoto, K. R. (1987). Two signals mediate hormone dependent nuclear localization of the glucocorticoid receptor. EMBO J. *6*, 3333–3340.

Roth, S., Stein, D., and Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the *dorsal* protein determines dorsoventral pattern in the Drosophila embryo. Cell *59*, 1189–1202.

Rushlow, C. A., Han, K., Manley, J. L., and Levine, M. (1989). The graded distribution of the *dorsal* morphogen is initiated by selective nuclear transport in Drosophila. Cell *59*, 1165–1177.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035–1048.

Schlissel, M. S., and Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell *58*, 1001–1007.

Sen, R., and Baltimore, D. (1986a). Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. Cell *47*, 921–928.

Sen, R., and Baltimore, D. (1986b). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell *46*, 705–716.

Simek, S., and Rice, N. R. (1988a). p59$^{v\text{-}rel}$, the transforming protein of reticuloendotheliosis virus, is complexed with at least four other proteins in transformed chicken lymphoid cells. J. Virol. *62*, 4730–4736.

Simek, S., and Rice, N. R. (1988b). Detection and characterization of the protein encoded by the chicken c-*rel* proto-oncogene. Oncogene Res. *2*, 103–119.

Singh, H., LeBowitz, J. H., Baldwin, A. S., and Sharp, P. (1988). Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site probe. Cell *52*, 415–423.

Stephens, R. M., Rice, N. R., Hiebsch, R. R., Bose, H. R., and Gilden, R. V. (1983). Nucleotide sequence of v-*rel*: the oncogene of the reticuloendotheliosis virus. Proc. Natl. Acad. Sci. USA *80*, 6229–6232.

Steward, R. (1987). *Dorsal*, an embryonic polarity gene in *Drosophila* is homologous to the vertebrate proto-oncogene, c-*rel*. Science *238*, 692–694.

Steward, R. (1989). Relocalization of the *dorsal* protein from the cytoplasm to the nucleus correlates with its function. Cell *59*, 1179–1188.

Steward, R., and Nüsslein-Volhard, C. (1986). The genetics of the *dorsal–bicaudal D* region of *Drosophila melanogaster*. Genetics *113*, 665–678.

Sylla, B. S., and Temin, H. M. (1986). Activation of oncogenicity of the c-*rel* proto-oncogene. Mol. Cell. Biol. *6*, 4709–4716.

Tempst, P., and Riviere, L. (1989). Examination of automated polypeptide sequencing using standard phenyl isothiocyanate reagent and subpicomole high performance liquid chromatographic analysis. Anal. Biochem. *183*, 290–300.

Tempst, P., Link, A. J., Riviere, L. R., Fleming, M., and Elicone, C. (1990). Internal sequence analysis of proteins separated on polyacrylamide gels at the sub-microgram level: improved methods, applications and gene cloning strategies. Electrophoresis, in press.

Walro, D. S., Herzog, N. K., Zhang, J., Lim, M. Y., and Bose, H. R. (1987). The transforming protein of avian reticuloendotheliosis virus is a soluble cytoplasmic protein which is associated with a protein kinase activity. Virology *160*, 433–444.

Wilhelmsen, K. C., Eggleton, K., and Temin, H. M. (1984). Nucleic acid sequences of the oncogene v-*rel* in reticuloendotheliosis virus strain T and its cellular homolog, the proto-oncogene c-*rel*. J. Virol. *52*, 172–182.

GenBank Accession Number

The accession number for the sequence reported in this paper is M37732.

Note Added in Proof

In Figure 3, the p50 sequence is also identical with *dorsal* at position 102.