Minireview

The structure of the NF-κB p50:DNA-complex: a starting point for analyzing the Rel family

Christoph W. Müller, Stephen C. Harrison*

Howard Hughes Medical Institute and Harvard University Department of Molecular and Cellular Biology, 7 Divinity Avenue, Cambridge, MA 02138, USA

Received 10 May 1995

Abstract The Rel family comprises a group of structurally related, eukaryotic transcription factors. The similarity extends over about 300 amino acid residues, the Rel homology region, which is responsible for DNA binding and dimerization. Two independently determined structures of homodimeric NF-κB p50 bound to DNA show the Rel homology regions and the DNA target sites. The protein consists of two β-barrel domains connected by a short linker. Five loops per monomer contact the DNA. Different half-site spacings in the two structures lead to different relative orientations of N- and C-terminal domains.

Key words: NF-κB; Rel family; DNA-binding protein; X-ray structure; Transcription

1. Introduction

NF-κB is a mammalian transcription factor that controls a number of genes important for immunity and inflammation. Examples include the genes for the Ig-κ light chain, the T-cell receptor α and β chains, MHC class I proteins, and cytokines such as GM-CSF, IL-6, IL-2, and TNF-α [1]. Viruses such as HIV-1 [2] use NF-κB for activating transcription of their own genes. The role of NF-κB in HIV infection and its importance in inflammatory processes makes it a potential target for drugs.

NF-κB consists of two subunits, p50 and p65 (RelA). Both subunits are part of a larger group of transcription factors: the Rel family [3,4]. In vertebrates five members are currently known: p50, p52, p65, RelB and c-Rel. Most of them are able to homo- and heterodimerize resulting in complexes of distinct DNA binding specificity with different cellular functions. Closely related to c-Rel is the oncoprotein v-Rel found in the retrovirus Rev-T (reviewed in [5]). Further family members are the *Drosophila* proteins Dorsal and Dif. Dorsal is a morphogen responsible for dorsal/ventral pattern formation in early development [6], while Dif controls a primitive cellular defense system in insect cells [7].

Common to all members of the Rel-family is a conserved stretch of about 300 amino acid residues, which is known as the Rel homology region (RHR). Within this region pairwise sequence identity varies around 50%. The entire RHR is needed for DNA-binding, in contrast to many other DNA-binding proteins, which have relatively small DNA-binding elements [8]. Dimerization requires only the C-terminal part of the RHR. Most family members form homo- or hetero-dimers both as free proteins and when bound to DNA. They bind very tightly to their DNA sites, with dissociation constants in the range of 10^{-12} M [1]. Well-known NF-κB heterodimer binding sites are found in the Ig κ light chain enhancer (GGGACTTTCC) [9] and in the interferon-β promoter site PRDII (GGGAAATTCC) [10]. A typical Dorsal binding site contains the sequence GGGAAAACCA, while the p50 homodimer prefers a sequence such as the one in the MHC class I enhancer (GGGGAATCCCC).

P50 and p52 contain only small extensions N- and C-terminal to the RHR. In vivo they are produced as precursor proteins, p105 and p100, respectively, which are proteolytically cleaved to their final size [11,12]. The longer variants p65, RelB, c-Rel, Dorsal and Dif contain C-terminal domains of variable length (Fig. 1a). These C-terminal extensions mediate the transcriptional activity once the respective homo- or heterodimers are bound to DNA [1,3,4]. In general only dimers containing at least one longer variant of the Rel family are able to activate transcription; p50 and p52 homodimers seem to act primarily as repressors.

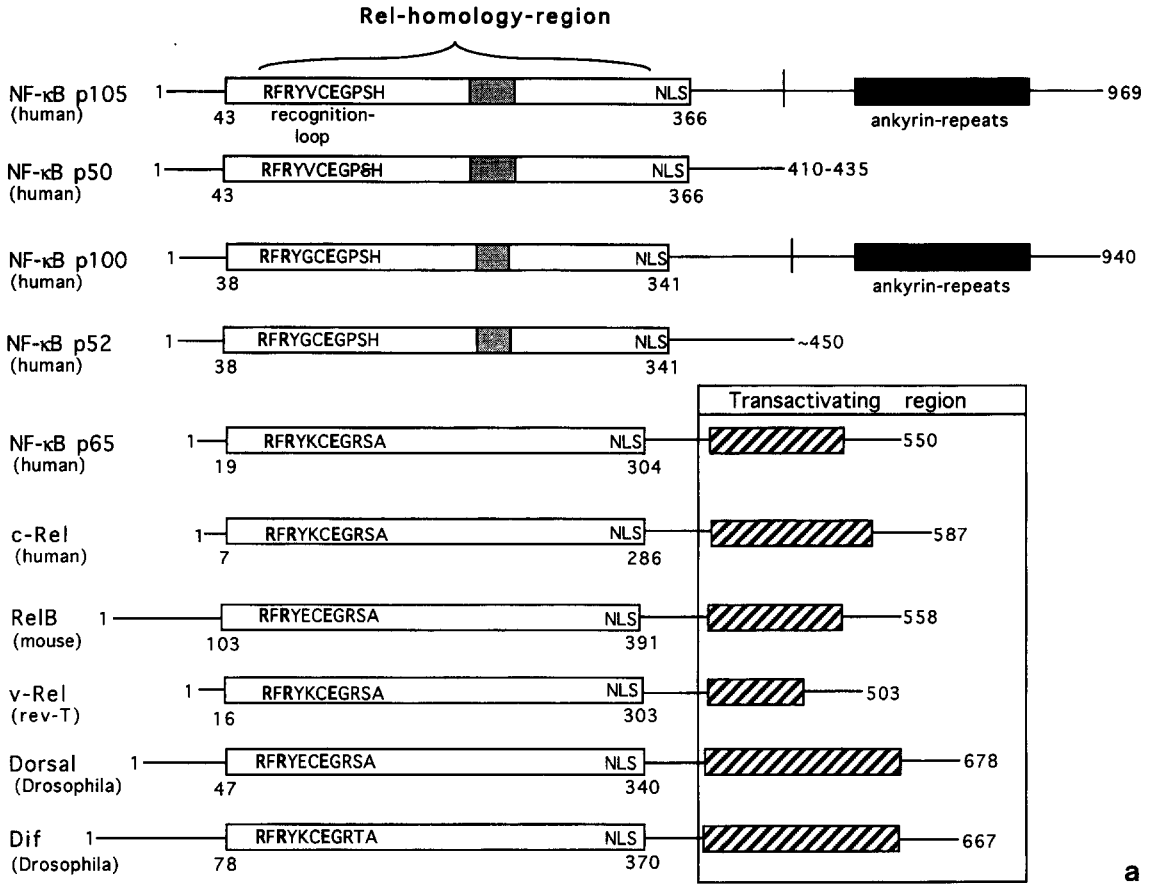
2. Cellular activation of NF-κB

NF-κB is activated by translocation from the cytosol to the nucleus. In the cytosol, the inhibitor molecule, IκB, masks the NF-κB nuclear localization sequence (NLS) [13]. Certain extracellular signals, such as cytokines, viruses and lipopolysaccharide (LPS), lead to phosphorylation of IκB by specific kinases. This event appears to render IκB susceptible to ubiquitination and subsequent proteolytic degradation by the proteosome [14]. The unmasked NLS regions direct NF-κB to the nucleus, where it can bind to its target site (Fig. 1b). Different species of IκB have been characterized [15]: IκBα and IκBβ both interact with the p50–p65 heterodimer. While IκBα is important for transient activation, IκBβ is involved in persistent long-term activation of NF-κB [16]. IκBγ corresponds to the C-terminal part of the p50 precursor p105. Further IκB homologs are the protooncogene Bcl3 [17] and the *Drosophila* homolog Cactus [18].

Numerous NF-κB-inducing extracellular effects have been described, but how these different signals lead to a common NF-κB activating event remains unclear. The interaction of different Rel-homologs in vivo is also incompletely understood.

*Corresponding author. Fax: (1) (617) 495 9613.

SSDI 0014-5793(95)00541-2



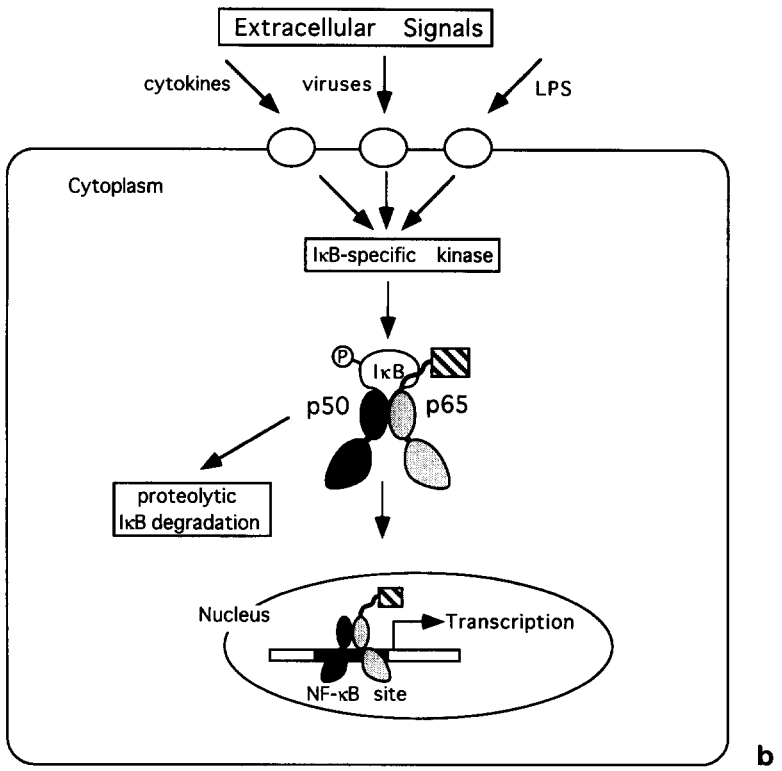


Fig. 1. (a) Overview of the Rel family. P50 and p52 are proteolytic cleavage products from their longer precursors p105 and p100. They do not activate trancription on their own. NLS, nuclear localization sequence. (b) Schematic diagram of the activation of NF-κB by translocation to the nucleus.

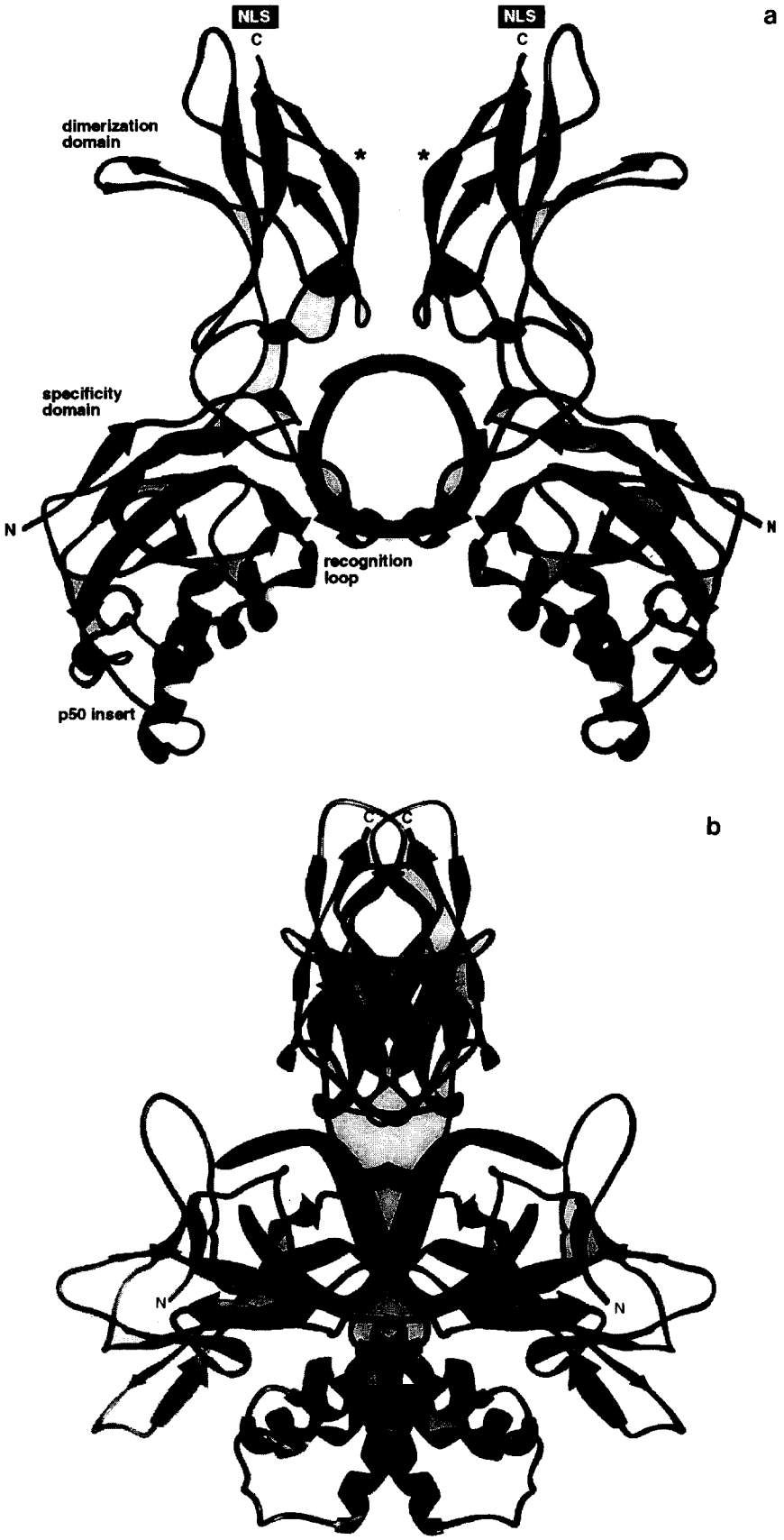


Fig. 2. Overall view of the p50 homodimer bound to DNA. NLS denotes the nuclear localisation sequence. Asteriks mark residue positions in a Dorsal mutant, which does not bind the IκB homolog Cactus. Fig. 2 was created with Ribbons [30]. (a) View along the DNA, with the dyad vertical. (b) View perpendicular to the DNA, the direction of the dyad is maintained.

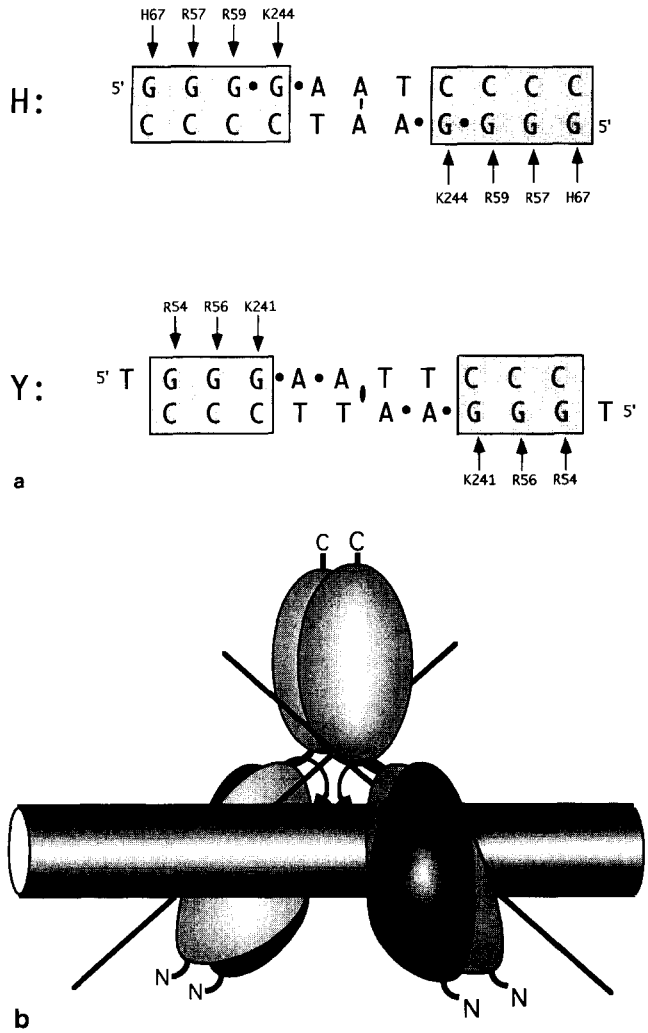


Fig. 3. (a) DNA target sites used for crystallization by the two groups. Important base specific contacts are listed. Heavy dots indicate the phosphates contacted in a conserved way by the dimerization domains. H: 11 bp duplex of the H structure corresponding to a MHC class I enhancer site; Y: 10 bp duplex of the Y-structure, corresponding to a symmetrized κB-site. (b) Schematic diagram of the domain movements of the p50 homodimer. The interface between the C-terminal domains and between C-terminal domains and DNA remains the same, but the phosphates contacted by the two subunits shift in register (see dots representing phosphates in 3a). Therefore in order to maintain base-specific contacts, the N-terminal domains rotate by 10–15 degrees about the indicated axes. H complex: light shading, Y complex: dark shading.

Recent inactivation of p50 and RelB genes in mice show the complexity of the system: Both types of mice develop to adulthood, but the inactivation of the gene for p50 results in increased vulnerability to acute infection, whereas RelB inactivation yields a complex pathological phenotype due to the absence of RelB in lymphoid cells [19,20].

3. Overview of the p50 complex

The crystal structure of an NF-κB p50 homodimer bound to DNA has been determined independently by two groups [21,22]. We refer to the two structures by the letters H and Y (denoting the institutions of the investigators). The structures were determined at 2.6 and 2.3 Å resolution, respectively, using very similar methods: a combination of multiple isomorphous replacement with iodine substituted nucleotides, anomalous diffraction from seleno-methionine substituted crystals, and molecular averaging. The H complex [21] contains a recombinant human p50 fragment (residues 2–366) and a 19 base oligonucleotide (5′-AGATGGGGAATCCCCTAGA-3′), which forms a central 11 bp duplex. The duplex corresponds to an MHC class I enhancer site, known to bind p50 homodimer [23][1]. The Y complex [22] contains a recombinant murine p50 fragment (residues 39–364) and a 11-base oligonucleotide (5′-TGGGAATTCCC-3′), which forms a 10 bp duplex. The duplex corresponds to a symmetrized consensus κB site [9]. A comparison of the two structures shows that different sequences of the DNA binding sites lead to different binding modes. The Rel homology region (RHR) comprises residues 43–366 (human) and 40–363 (mouse). Only residues 43–352 and 39–350, respectively are clearly defined in the two structures. Within the RHR, the sequences of human and murine p50 are almost identical. Thus the key difference between the H and Y structures is the halfsite spacing (Fig. 3).

The p50 RHR folds into two domains, connected by a short linker (Fig. 2). Both domains contain β-barrels related to the immunoglobulin (Ig) fold [24]. The 200-residue, N-terminal domain is based on an I-type Ig barrel [25], with a small three-stranded sheet and an α-helical subdomain added at one end. The helical subdomain contains the short 'insert', a segment found only in the RHR of p50 and p52 (Fig. 1a). The 100 residue C-terminal domain is a C-type Ig barrel.

The dimer wraps into the major groove, so that it nearly encloses the DNA. Loops connecting strands in the β-barrels of both domains contact DNA bases and backbone. The linker must be flexible in order to allow the dimer to open and close its N-terminal 'jaws' around the DNA.

The dimer interface is formed by the four-stranded sheet of the C-terminal domain. A core of hydrophobic residues surrounded by polar residues form an interface of 700 Å^2. This packing arrangement is unique to the Rel family and different from the packing of other Ig-like domains such as those in human growth factor receptors, antibodies, or MHC proteins. Most of the residues involved in forming the interface are strictly conserved within the Rel family. Residues in the C-terminal domain that contact DNA backbone are adjacent to the dimer interface, and the subunit:subunit and subunit:DNA contacts in this region form a continous recognition surface.

4. DNA recognition

Residues from five different inter-strand loops contact DNA. One of these loops is the interdomain linker, and there are two from each of the domains. Thus the entire RHR contributes to DNA binding. The protein:DNA interface has a large solvent-excluded area (2300 Å^2 per dimer), consistent with the unusually high affinity.

Base specific contacts in the H complex are restricted to residues from the N-terminal domain, which we consequently

[1] In the crystals used for the structure determination there is an A:A mismatch at the dyad; crystals grown using DNA with a central A:T basepair show that the A:A mismatch produces no significant distortion.

refer to as the 'specificity domain'. All but one of these residues are in the 'recognition loop', which connects strands A and B. The most important residues appear to be two arginines (57 and 59 in the H numbering scheme) and a glutamic acid (63). Each of the arginines contacts a guanine, and salt links to the glutamate create a network that holds the two arginines in place. These residues, which are conserved throughout the Rel family, form a unit for recognition of two successive G's (see Fig. 3a). Sequences preferred by p50, such as the MHC site, have 3 or 4 G's in each half-site. In the H-structure the outermost G of each halfsite is contacted by His67, and the innermost, by Lys244 (the last residue in the specificity domain). These residues are Ala and Arg, respectively, in p65, which appears to require only two guanines per half-site. In the Y-structure, two residues from the C-terminal domain extend sufficiently far into the major groove to contact bases; these same residues make backbone contacts in the H complex.

5. Comparison between the two structures

As expected, the two independently determined structures show identical protein folds. However the different half-site spacings (Fig. 3a) result in different relative orientations of the two domains. Taken separately, the N-terminal and C-terminal domains agree extremely well. Moreover, the paired dimerization domains and the most strongly contacted phosphates of both structures superpose closely. Relative to this superposition, the specificity domains have to rotate by 10–15 degrees to maintain their base-specific contacts. The hinge point lies close to the interdomain linker, and the axis of the observed rigid-body rotation of the N-terminal domain runs roughly through the points of contact that the domain makes with DNA backbone (Fig. 3b).

The DNA duplex adopts a similar conformation in both structures. There is a deep major groove, and the B-like helix is somewhat underwound. The DNA axis in the H complex bends by about 15 degrees to either side of the central AAT; the DNA in the Y complex is essentially straight. The difference in bending appears to be one of the ways in which the complex accomodates to the different half-site spacings.

6. Interactions with IκB, HMGI and DSP1

Transcription is regulated on numerous levels. The size and diversity of the surface of the NF-κB p50 homodimer suggests how its activity can be modulated through interactions with other factors. The structure shows a large groove above the dimer interface (Fig. 2a). Two lines of evidence suggest that IκB binds in this groove. First, the NLS lies just at its outer rim: an intact NLS is necessary for the IκB interaction [26]. Second, a mutant in Dorsal, that interferes with binding of Cactus (the *Drosophila* IκB homolog), maps to a site within the groove [27]. How IκB prevents DNA-binding remains unclear. To study this question, the crystal structure of an IκB:NF-κB complex seems an obvious goal.

Two molecules are known to modulate the transcriptional activity of NF-κB. The high mobility group protein HMGI(Y) is required for transcriptional activity of NF-κB bound to the IFNb promoter [28]. HMGI(Y) contacts the minor groove, which is exposed in our structure. Potential interaction surfaces with NF-κB are probably the helical extensions in the N-terminal domain. The Dorsal switch protein (DSP1) binds to region adjacent to the Dorsal/NF-κB binding sites and converts Dorsal and NF-κB to repressors [29]. A direct interaction with Dorsal and NF-κB is observed, but precise identification of the interaction surface with Rel family proteins is still to be determined.

References

[1] Baeuerle, P.A. and Henkel, T. (1994) Annu. Rev. Immunol. 12, 141–179.
[2] Ross, E.K., Buckler-White, A.J., Rabson, A.B., Englund, G. and Martin, M.A. (1991) J. Virol. 65, 4350–4358.
[3] Liou, H.C. and Baltimore, D. (1993) Curr. Op. Cell Biol. 5, 477–487.
[4] Blank, V., Kourilsky, P. and Israel, A. (1992) Trends Biochem. Sci. 17, 135–140.
[5] Gilmore, T.D. (1991) Trends Genet. 7, 318–322.
[6] Steward, R. (1987) Science 238, 692–694.
[7] Ip, Y.T., Reach, M., Engstrom, Y., Kadalayil, L., Cai, H., Gonzalez-Crespo, S., Tatei, K. and Levine, M. (1993) Cell 75, 753–763.
[8] Harrison, S.C. (1991) Nature 353, 715–719.
[9] Sen, R. and Baltimore, D. (1986) Cell 46, 705–716.
[10] Lenardo, M.J., Fan, C.-M., Maniatis, T. and Baltimore, D. (1989) Cell 57, 287–294.
[11] Rice, N.R., MacKichan, M.L. and Israel, A. (1992) Cell 71, 243–253.
[12] Mercurio, F., DiDonato, J.A., Rosette, C. and Karin, M. (1993) Genes Dev. 7, 705–18.
[13] Beg, A.A., Ruben, S.M., R.I., S., Haskill, S., Rosen, C.A. and Baldwin, A.S. (1992) Genes Dev. 6, 1899–1913.
[14] Palombella, V.J., Rando, O.J., Goldberg, A.L. and Maniatis, T. (1994) Cell 78, 773–785.
[15] Gilmore, T.D. and Morin, P.J. (1993) Trends Genet. 9, 427–433.
[16] Thompson, J.E., Phillips, R.J., Erdjument-Bromage, H., Tempst, P. and Ghosh, S. (1995) Cell 80, 573–582.
[17] Ohno, H., Takimoto, G. and McKeithan, T.W. (1990) Cell 60, 991–997.
[18] Kidd, S. (1992) Cell 71, 623–635.
[19] Sha, W.C., Liou, H.-C., Tuomanen, E.I. and Baltimore, D. (1995) Cell 80, 321–330.
[20] Weih, F., Carrasco, D., Durham, S.K., Barton, D.S., Rizzo, C.A., Ryseck, R.P., Lira, S.A. and Bravo, R. (1995) Cell 80, 331–340.
[21] Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L. and Harrison, S.C. (1995) Nature 373, 311–317.
[22] Ghosh, G., Van Duyne, G., Ghosh, S. and Sigler, P.B. (1995) Nature 373, 303–310.
[23] Baldwin, A.S. and Sharp, P.A. (1987) Mol. Cell. Biol. 7, 305–313.
[24] Bork, P., Holm, L. and Sander, C. (1994) J. Mol. Biol. 242, 309–320.
[25] Harpaz, Y. and Chothia, C. (1994) J. Mol. Biol. 238, 528–539.
[26] Beg, A.A. and Baldwin, A.S. (1993) Genes Dev. 7, 2064–2070.
[27] Lehming, N. and Ptashne, M. (1995) Proc. Natl. Acad. Sci. USA, in press.
[28] Thanos, D. and Maniatis, T. (1992) Cell 71, 777–789.
[29] Lehming, N., Thanos, D., Brickman, J.M., Ma, J., Maniatis, T. and Ptashne, M. (1994) Nature 371, 175–179.
[30] Carson, M. (1987) J. Mol. Graphics 5, 103–106.