Rel revealed: cocrystal structures of the NF-κB p50 homodimer

Crystal structures of the NF-κB p50 homodimer, a member of the Rel family of transcription factors, bound to a DNA target give insight into how these proteins use loops between β-strands to recognize DNA and how they are inactivated by IκB.

Chemistry & Biology February 1995, **2**:77-81

NF-κB is a prototypic member of a large family of eukaryotic transcription factors that share a 300 amino acid region of homology to the product of the *rel* oncogene (Fig. 1). The Rel homology region is responsible for both DNA binding and dimerization; Rel family members must form either homodimers or heterodimers to function. The Rel family members differ from most transcriptionally-active, DNA-binding proteins in two important respects. First, most DNA-binding proteins have a small subdomain or module within the protein that can interact with DNA independently; in Rel family members, in contrast, the complete Rel homology region is required for DNA recognition. Second, unlike other DNA-binding proteins, NF-κB activity is controlled by interactions with other, inhibitory, proteins which restrict its access to the nucleus, keeping it in the cytoplasm (for example, IκB sequesters the biologically-active NF-κB p50–p65 heterodimer). A most impressive combination of structural and molecular biology has now explained these unique features of the Rel proteins, proving yet again that there is often no substitute for a direct look at a biological system, and setting the stage for detailed, atomic-level dissection of the mechanism of action of Rel proteins. Sigler and coworkers [1] and Harrison and coworkers [2] have determined two highly-similar and entirely novel three-dimensional structures of the NF-κB p50 homodimer bound to distinct DNA targets. Recent studies using the *Drosophila* proteins Dorsal (a homolog of NF-κB) and Cactus (an IκB homolog) have also provided insight into the mechanism of Rel protein control *in vivo* (N. Lehming, *et al.*, personal communication).

Rel proteins as transcription factors

NF-κB was first purified from B lymphocytes on the basis of its ability to bind to a ten-base-pair κB site in the immunoglobulin κ light-chain enhancer element [3]. Further work demonstrated that a heterodimer of the p50 and p65 (RelA) subunits is required for DNA-binding activity. The p50–p65 heterodimer binds directly to DNA, and is responsible for controlling B-cell development. It also activates the expression of specific host-defense genes in response to infection and other forms of stress in all cell types (reviewed in [4]). Other Rel family members (such as Dorsal from *Drosophila*) affect transcription as homodimers [5], and homodimers of both p50 and the related p52 protein

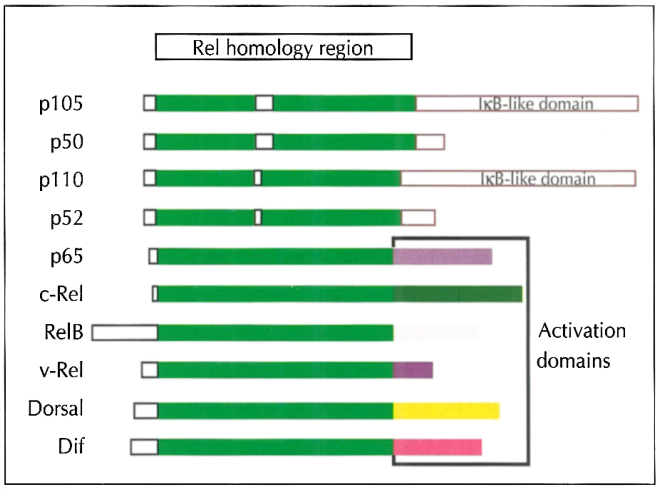

Fig. 1. Schematic diagram representing a primary sequence alignment of members of the Rel/NF-κB family. The green segments denote regions homologous to the product of the Rel oncogene, which are both necessary and sufficient for dimerization and DNA binding. A ubiquitin–proteasome conjugated pathway is responsible for proteolysis of the carboxy-terminal IκB-like domains of p105 and p110 to give p50 and p52, respectively. The p65 (RelA), c-Rel, RelB, v-Rel, Dorsal and Dif proteins are composed of Rel homology regions and non-homologous transcriptional activation domains (indicated with different colors). Reprinted with permission from [1].

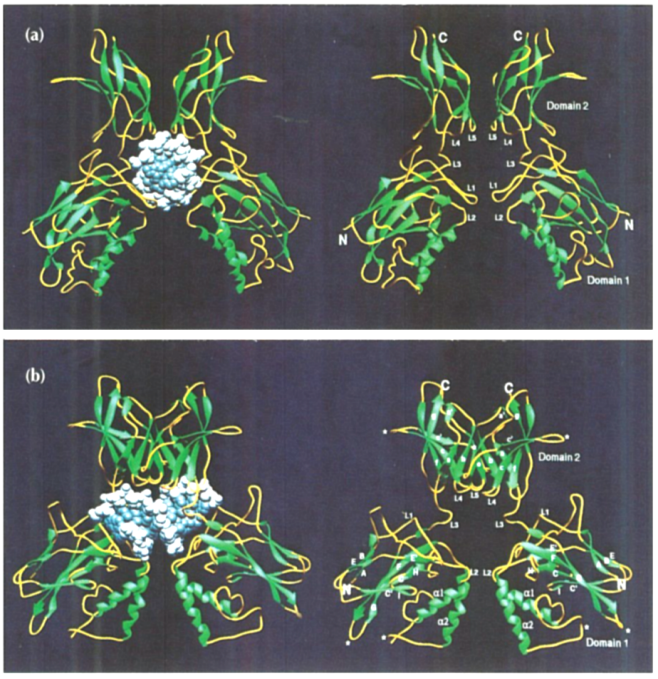

Fig. 2. Structure of the homodimeric NF-κB p50 subunit bound to an oligonucleotide containing a κB recognition element. **(a)** Ribbon drawings of the entire complex viewed down the DNA helix axis with the dyad axis running vertical (left). The same view with the van der Waals space filling model of the oligonucleotide removed and the DNA interacting loops labeled (right). **(b)** Figure in (a) rotated 40° around the vertical dyad. Secondary structural elements are labeled (right). Reprinted with permission from [1].

can in fact compete with NF-κB for its DNA targets, modulating the transcription of certain genes (reviewed in [6]).

NF-κB subunits have been found in all mammalian cells, but are typically inactive because they are limited to the cytoplasm by inhibitors of the IκB family. These inhibitory proteins contain repeats of the ankyrin motif, which are necessary for binding to NF-κB. IκB binding to NF-κB effectively masks its nuclear localization signal, preventing transit into the nucleus (reviewed in [7]). Viral infections and other inducers of NF-κB, such as lipopolysaccharide and phorbol ester treatment, trigger dissociation and/or degradation of IκB, allowing the p50/p65 heterodimer to enter the nucleus and activate transcription in cell types other than B lymphocytes (reviewed in [4,6]). Paradoxically, human immunodeficiency viruses (HIV) and some other viral pathogens exploit the very mechanism that regulates transcription of host defense genes. Reverse transcription of the HIV-1 RNA genome creates a pair of DNA-binding sites for NF-κB, which in turn activates transcription of the gene encoding the retroviral polyprotein, allowing viral replication.

Structure determination

Two highly-similar structures were obtained with truncated forms of p50. The p50 forms used by the two groups had different amino-termini, but nearly identical carboxyl-termini, and included the carboxy-terminal nuclear localization signal. These proteolytically-defined p50 fragments (amino acids 39–364 [1] and 2–366 [2]) form symmetric homodimers and bind specifically to DNA sites recognized by NF-κB. Ghosh *et al.* [1] used a minimal 10-base-pair palindromic κB site with an overhanging 'T' at the 5' end for cocrystallization. In contrast, Muller *et al.* [2] cocrystallized their p50 construct with a very unusual 19-base-pair oligonucleotide, containing a quasi-palindromic 11-base-pair site (with a single A:A mismatch at the center) flanked by four bases at either end which made contacts with neighboring protein–DNA complexes in the crystal lattice. Both structure determinations were challenging enough to require the entire modern crystallographic armamentarium, including data collection with extremely intense synchrotron wiggler beam lines and Fuji image plates at liquid nitrogen temperature, a combination of multiple isomorphous replacement and multiple wavelength anomalous dispersion to obtain initial experimental phases, and considerable computational manipulation of the phase information to obtain interpretable electron-density maps. Technical difficulties notwithstanding, the final resolution limits of 2.3 and 2.6 Å, respectively, and the high quality of the refined structures are a testament to the impressive power of modern protein crystallography.

Dimerization and DNA binding

The three-dimensional structure of the symmetric p50 homodimer–DNA complex resembles a butterfly with the protein domains as wings surrounding a cylindrical B-form DNA (Fig. 2). Each polypeptide chain folds into

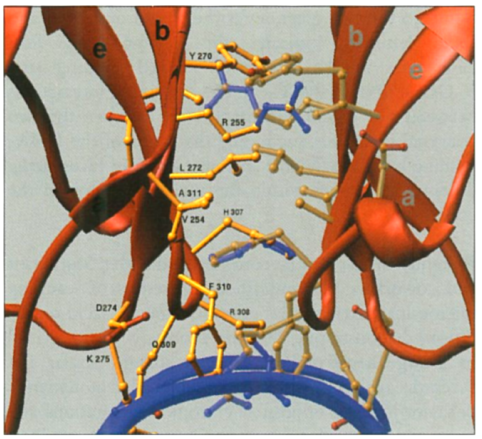

Fig. 3. The dimer interface, viewed along the DNA helix axis, showing side chains mediating dimerization and DNA contacts. Reprinted with permission from [2].

two domains, and it is immediately apparent that there is no DNA-binding subdomain within the Rel homology region. Residues from the entire length of the protein contribute to the DNA-interaction surface.

Both domains contain β-sandwich core structures with immunoglobulin-like folds, which is a remarkable coincidence when one considers that NF-κB regulates transcription of genes encoding immunoglobulins. The larger amino-terminal domain is connected to its carboxy-terminal partner by a loop region that is well defined in the electron-density map and contributes to interactions with the DNA. In addition to its immunoglobulin-like core structure, domain 1 contains an insertion of two α-helices, which pack in an anti-parallel fashion against the medial β-sheet face without interacting with the other subunit. Dimerization is supported entirely by the smaller immunoglobulin-like carboxy-terminal domain, and results from a series of interdigitating hydrophobic residues projecting away from the medial β-sheets of the two p50 subunits (Fig. 3). Sequence alignments demonstrate that the amino acids responsible for dimerization and those defining the hydrophobic core of the carboxy-terminal domain are highly conserved, suggesting that the p50 homodimer structure is a good model for both homo- and heterodimerization of other Rel family members. The biological importance of protein dimerization was the subject of an earlier review in this journal [8].

Figs 2 and 4 reveal a mode of DNA binding for which there is no precedent (for recent reviews of more typical DNA-binding motifs see [9–11]). The p50 homodimer embraces a slightly bent, slightly unwound, B-form double helix, donating a total of ten loop regions for DNA contacts with both

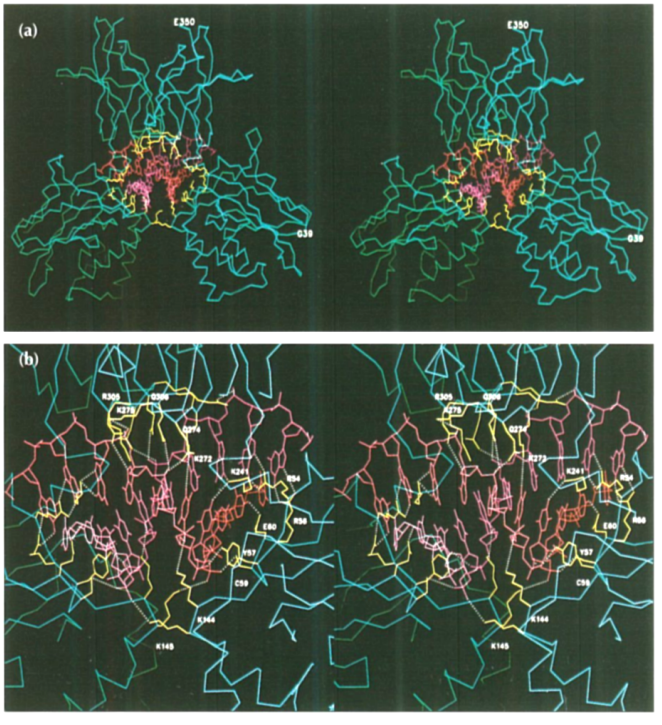

Fig. 4. Stereoviews of protein–DNA interactions in homodimeric NF-κB p50 bound to an oligonucleotide containing a κB recognition element. **(a)** α-Carbon trace (blue) of both p50 subunits and complete atomic stick figure model of the κB element (red and purple) displaying side chain and backbone contacts (yellow). **(b)** Expanded image of the central region in (a) showing hydrogen bonds (white dotted lines). Reprinted with permission from [1].

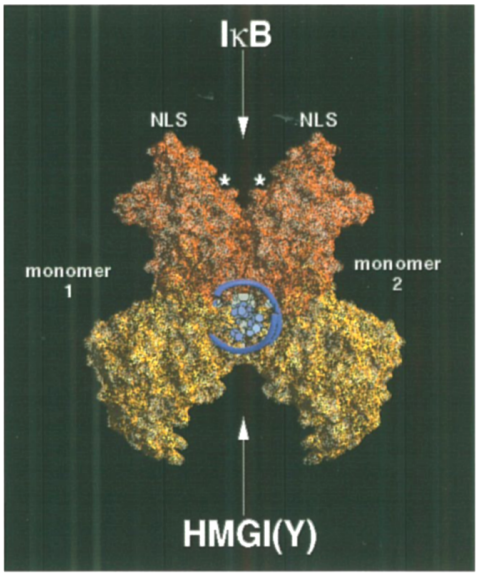

Fig. 5. Surface representation of the NF-κB p50 homodimer bound to DNA, showing likely regions for interactions with IκB and the high-mobility group protein I(Y) (HMG I(Y)), which is essential for transcriptional activation at some NF-κB sites. Asterisks mark the positions of two Dorsal mutations, which inhibit binding of the IκB homolog Cactus. Rel family members also interact with non-IκB proteins during transcription (reviewed in [6]). For example, NF-κB and HMG I(Y) simultaneously occupy complementary portions of the human interferon-β promoter [17]. Reprinted with permission from [2]. NLS, nuclear localization signal.

phosphodiester/ribose backbones and the major groove edges of some bases. The extensive protein–nucleic acid interaction surface (the buried solvent accessible area exceeds 3500 Å^2) includes the entire κB site, and each subunit makes contacts throughout the length of the recognition element. As in many protein–DNA complexes, both direct and water-mediated contacts are seen within the interface.

What do the p50 homodimer cocrystal structures tell us about Rel DNA-binding specificity? The NF-κB p50/p65 heterodimer binds with extremely high affinity (K_d ~10^{-12} M) to both symmetric and asymmetric decameric DNA sequences of consensus 5'-GGGPuNNPyPyCC-3' [12], and site-selection studies with p50 and p65 homodimers suggest that these subunits have quite distinct DNA-binding specificities [13]. *In vivo*, Rel homo- and heterodimers modulate transcription of a wide range of cellular and viral genes via interactions with a similarly varied collection of target sites. Although the two cocrystal structures were obtained with recognition elements of different length and sequence, there are similar features that may be common to other Rel family members. In each p50 subunit, two Arg side chains donate a pair of hydrogen bonds to the two guanines at the 5' end of each strand (bold lettering indicates the corresponding four base pairs in the consensus NF-κB binding site, 5'-**GG**GPuNNPyPy**CC**-3'). The remaining interactions between side chains and bases differ between the two cocrystal structures, suggesting that some of the DNA-binding loops are flexible and are capable of recognizing different sequences within the variable portion of the consensus site.

Recently, Spolar and Record [14] and Sigler, Sturtevant and coworkers [15] published the results of scanning microcalorimetric studies of protein–nucleic acid recognition, demonstrating that the process is characterized by a large negative heat capacity change. Spolar and Record analyzed this remarkable thermodynamic behavior in terms of possible entropic contributions, and suggested that molecular recognition proceeds via local folding or 'induced fit' [16] of the protein and/or nucleic acid components. Sequence-specific DNA binding via flexible loops by members of the Rel family of transcription factors could provide useful examples with which to test this 'induced fit' model of protein–nucleic acid recognition.

Mechanism of Rel inhibition

Rel inhibition by IκB effectively masks the Rel nuclear localization signal (NLS), which is disordered in both cocrystal structures. Fig. 5 illustrates the position of the first residue of the NLS and demonstrates its proximity to the sites of two mutations in Dorsal (the *Drosophila* homolog of NF-κB) that prevent interaction with Cactus (the *Drosophila* homolog of IκB) (N. Lehming, *et al.*, personal communication). In addition, a mutation in Cactus suppresses the effects of both Dorsal changes. These data suggest that IκB interacts with both the NLS and the sites on p50 corresponding to the Dorsal mutations (Fig. 5). Although it is not yet clear how IκB masks the NLS, these structural models provide an immensely detailed foundation from which to design experiments to ask this question.

References

1. Ghosh, G., Duyne, G.V., Ghosh, S. & Sigler, P.B. (1995). Structure of NF-κB p50 homodimer bound to a κB site. *Nature* **373**, 303–310.
2. Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L. & Harrison, S.C. (1995). Structure of the NF-κB p50 homodimer bound to DNA. *Nature* **373**, 311–317.
3. Sen, R. & Baltimore, D. (1986). Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. *Cell* **47**, 921–928.
4. Baeuerle, P. & Henkel, T. (1994). Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.* **12**, 141–179.
5. Ip, Y., Park, R., Yazdanbakhsh, K. & Levine, M. (1992). The Dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1518–1530.
6. Liou, H.-C. & Baltimore, D. (1993). Regulation of the NF-κB/Rel transcription factor and IκB inhibitor system. *Curr. Opin. Cell. Biol.* **5**, 477–487.
7. Nolan, G. & Baltimore, D. (1992). The inhibitory ankyrin and activator Rel proteins. *Curr. Opin. Genet. Dev.* **2**, 211–220.
8. Austin, D., Crabtree, G. & Schreiber, S. (1994). Proximity versus allostery: the role of regulated protein dimerization in biology. *Chemistry & Biology* **1**, 131–136.

9. Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.
10. Wolberger, C. (1993). Transcription factor structure and DNA binding. *Curr. Opin. Struct. Biol.* **3**, 3–10.
11. Burley, S.K. (1994). DNA-binding motifs from eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **4**, 3–11.
12. Grimm, S. & Baeuerle, P. (1993). The inducible transcription factor NF-κB: structure–function relationship of its protein subunits. *Biochem. J.* **290**, 297–308.
13. Kunsch, C., Ruben, S. & Rosen, C. (1992). Selection of optimal κB/Rel DNA-binding motifs: interaction with both subunits of NF-κB with DNA is required for transcriptional activation. *Mol. Cell. Biol.* **12**, 4412–4421.
14. Spolar, R. & Record, M. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777–784.
15. Ladbury, J., Wright, J., Sturtevant, J. & Sigler, P. (1994). A thermodynamic study of the *trp* repressor–operator interaction. *J. Mol. Biol.* **238**, 669–681.
16. Koshland, D. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* **44**, 98–104.
17. Thanos, D. & Maniatis, T. (1992). The high mobility group protein HMG I(Y) is required for NF-κB-dependent virus induction of the human IFN-β gene. *Cell* **71**, 777–789.

Stephen K Burley, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.