The EMBO Journal Vol.16 No.23 pp.7078–7090, 1997

Structure of the human NF-κB p52 homodimer–DNA complex at 2.1 Å resolution

Patrick Cramer[1], Christopher J.Larson[2,3], Gregory L.Verdine[2] and Christoph W.Müller[1,4]

[1]European Molecular Biology Laboratory (EMBL), Grenoble Outstation, c/o ILL, BP 156, 38042 Grenoble Cedex 9, France, [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA, [3]The Salk Institute for Biological Studies, PO Box 85800, San Diego, CA 92186-5800, USA

[4]Corresponding author
e-mail: mueller@embl-grenoble.fr

The crystal structure of human NF-κB p52 in its specific complex with the natural κB DNA binding site MHC H-2 has been solved at 2.1 Å resolution. Whereas the overall structure resembles that of the NF-κB p50–DNA complex, pronounced differences are observed within the 'insert region'. This sequence segment differs in length between different Rel proteins. Compared with NF-κB p50, the compact α-helical insert region element is rotated away from the core of the N-terminal domain, opening up a mainly polar cleft. The insert region presents potential interaction surfaces to other proteins. The high resolution of the structure reveals many water molecules which mediate interactions in the protein–DNA interface. Additional complexity in Rel protein–DNA interaction comes from an extended interfacial water cavity that connects residues at the edge of the dimer interface to the central DNA bases. The observed water network might acount for differences in binding specificity between NF-κB p52 and NF-κB p50 homodimers.

Keywords: crystal structure/NF-κB p52/protein–DNA interaction/Rel protein/transcription

Introduction

The NF-κB/Rel family of eukaryotic transcription factors controls many mammalian genes of significant biomedical importance, including genes encoding pro-inflammatory cytokines, interferones, major histocompatibility complex (MHC) proteins, growth factors, cell adhesion molecules, but also viruses such as the human immunodeficiency virus (HIV) or Herpes (Baeuerle and Henkel, 1994; Thanos and Maniatis, 1995; Baeuerle and Baltimore, 1996; Baldwin, 1996; Chytil and Verdine, 1996). Members of the NF-κB/Rel family can be divided into two subgroups. The first subgroup includes the mammalian proteins p65, RelB and c-Rel as well as the *Drosophila* proteins Dorsal, Dif and Relish, and the recently discovered protein Gambif1 from *Anopheles gambiae* (Barillas-Mury *et al.*, 1996). The insect proteins are involved in morphogenesis (Dorsal) and primitive immune defense (Dif, Relish and Gambif1). In addition to the Rel homology region (RHR), all members of this subgroup have putative transcriptional activation domains. The second subgroup includes the mammalian precursor proteins p105 and p100 which are proteolytically cleaved *in vivo* into the mature proteins p50 and p52, respectively (Baeuerle and Henkel, 1994). These proteins lack an activation domain. Most of the NF-κB/Rel family members are able to form homo- and heterodimers of distinct DNA binding specificity. Activity and cellular localization of NF-κB/Rel proteins are further regulated through interactions with proteins of the IκB family (Gilmore and Morin, 1993).

The NF-κB subunits p50 and p52 share 63% identical amino acid residues in the RHR (Figure 1A). They show functional similarities. Both proteins can form transactivating heterodimeric complexes with p65 or c-Rel (Lin *et al.*, 1995) and bind with comparable affinities to NF-κB p65 (Schmid *et al.*, 1991). Their homodimers are both thought to repress transcription via competition with transactivating complexes for binding sites. However, NF-κB p50 and p52 are distinguishable by their cell-type distribution (Baldwin and Sharp, 1987, 1988) and also show some functional differences. NF-κB p50 homodimers bind equally well (dissociation constants in the low picomolar range) to the pseudo-symmetric MHC H-2 site (5′-GGGGATTCCCC-3′) and the non-symmetric Ig/HIV site (5′-GGGGACTTTCC-3′). In contrast, NF-κB p52 homodimers show comparable affinity to the MHC H-2 site but their affinity to the Ig/HIV site is ~15-fold lower than that of NF-κB p50 (Schmid *et al.*, 1994). The heterodimers p50/p65 and p52/p65 both bind tightly to the Ig/HIV site (Perkins *et al.*, 1992; Duckett *et al.*, 1993). Furthermore, NF-κB p50 and p52 interact differently with Bcl-3, a member of the IκB family (Gilmore and Morin, 1993). Whereas Bcl-3 can remove p50 homodimers from DNA (Franzoso *et al.*, 1993), it associates with p52 homodimers through its ankyrin domain and forms a ternary complex with DNA at κB sites allowing for transcriptional activation through its activation domains (Bours *et al.*, 1993).

NF-κB p50–DNA complex structures of the mouse homologue at 2.3 Å resolution (Ghosh *et al.*, 1995) and of the human homologue at 2.6 Å resolution (Müller *et al.*, 1995) have been reported. Here we present the crystal structure of the human NF-κB p52 homodimer in its specific complex with DNA at 2.1 Å resolution. The structure provides information on a second Rel family member and gives further insight into DNA recognition by Rel proteins.

Results and discussion

Overall structure

The Rel homology region (RHR) of human NF-κB p52 was co-crystallized with a DNA duplex comprising an MHC H-2 site (Figure 1B), which is the preferred binding

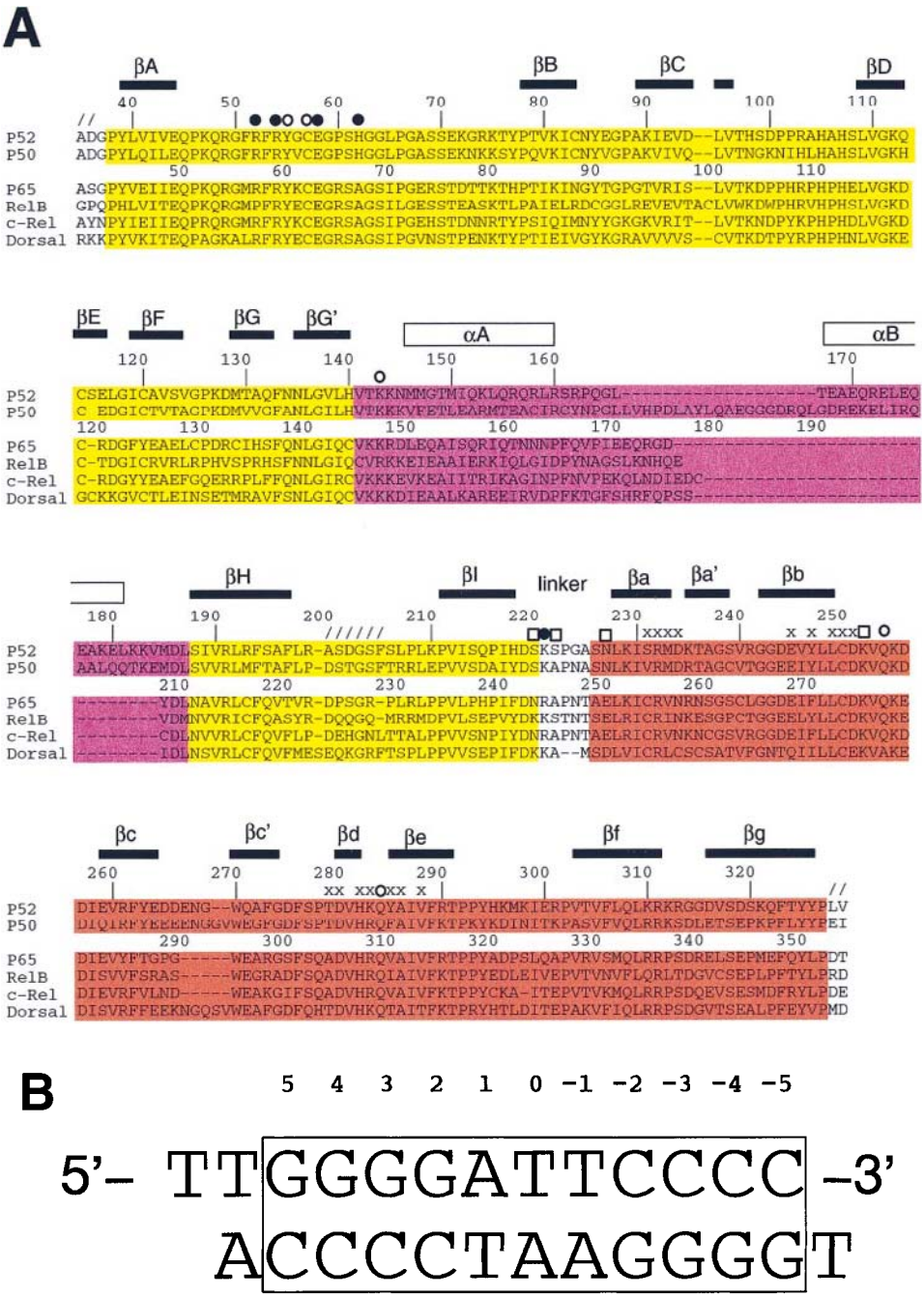

Fig. 1. (**A**) Amino acid sequence alignment of human NF-κB p52, p50, p65, RelB and c-Rel and the *Drosophila* protein, Dorsal. Amino acid numbering for human NF-κB p52 and p50 is given above and below the sequence, respectively. Residues of the N-terminal domain core (residues 38–140 and 188–220), the insert region (residues 141–187) and the C-terminal domain (residues 226–327) are colored in yellow, magenta and red, respectively. Secondary structure elements are indicated as lines (β-strands) and boxes (α-helices). The secondary structure prediction program PHD (Rost and Sander, 1993; Rost, 1996) predicts the presence of the insert region helix αA in p65 and Dorsal. Therefore, residues within the insert region have been aligned assuming conservation of this helix in all family members. Hatched bars mark residues which are disordered in the X-ray structure. Black crosses mark residues in the dimer interface. Filled and open circles indicate residues involved in direct contacts to DNA bases and the DNA backbone, respectively. Open squares indicate residues making water-mediated DNA contacts. (**B**) Sequence of the DNA duplex used in co-crystallization. Base pair numbering is indicated above the sequence. The box contains the natural 11 bp κB binding site MHC H-2 which is well ordered in our structure. The ends of the duplex (outside the box) are disordered.

site for NF-κB p52 homodimers (Schmid *et al.*, 1991). The same binding site is present in the human NF-κB p50 complex structure except for a central A:A mismatch. The NF-κB p52 dimer wraps around the DNA giving the complex the appearance of a butterfly (Figure 2A). Each monomer consists of two immunoglobulin-like domains connected by a short linker. The DNA is contacted by five loops per monomer, two of them in the N-terminal domain, two in the C-terminal domain and one being the interdomain linker. The N-terminal, nine-stranded β-barrel contains the 'recognition loop' which contacts three concomitant G:C bp in the major groove, while residues of the seven-stranded C-terminal β-barrel form the dimer interface. The overall structure shows strong similarity to

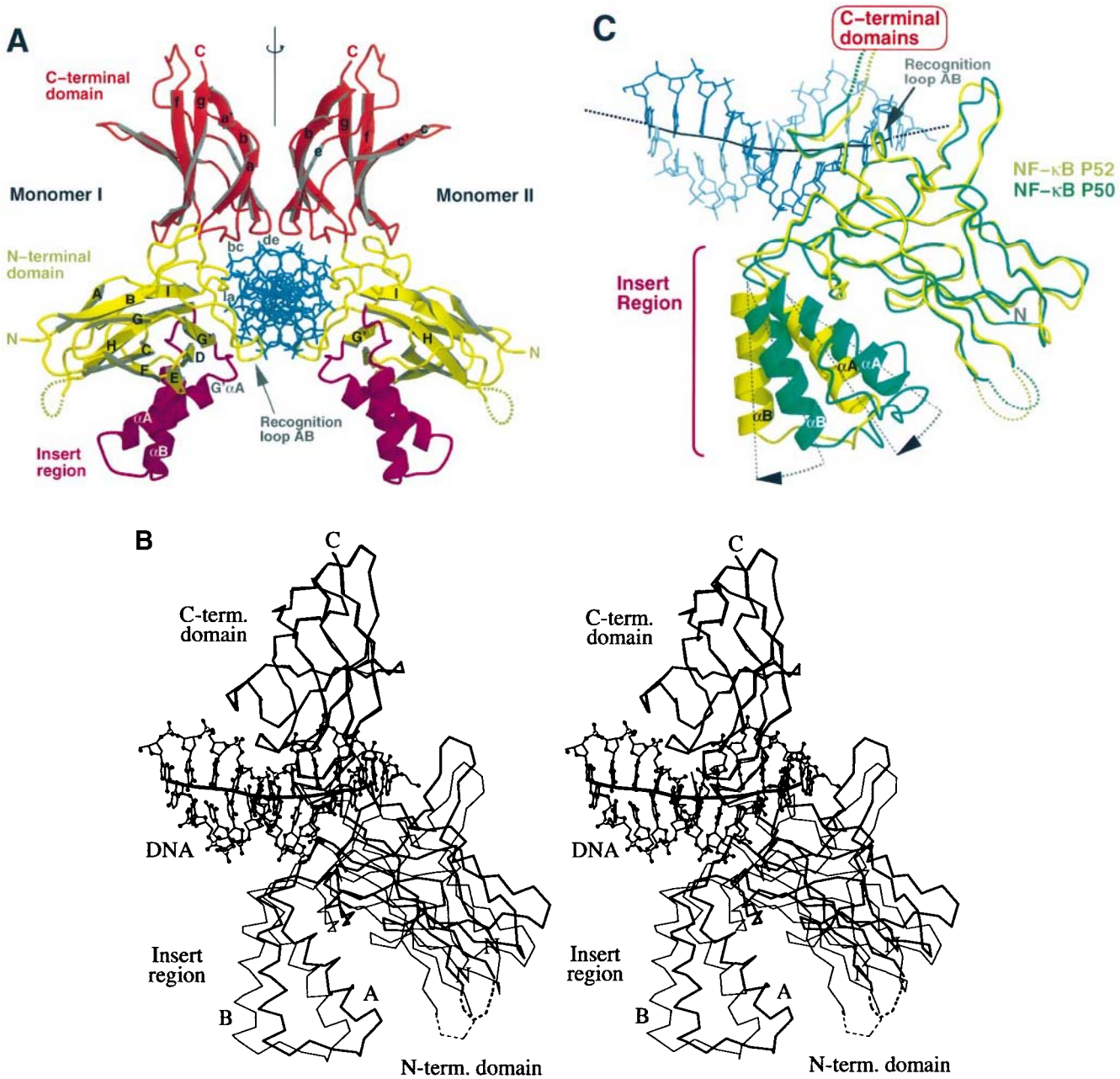

Fig. 2. (**A**) Overall view of the NF-κB p52 homodimer–DNA complex structure along the DNA helical axis with the approximate dyad vertical. The DNA duplex is in blue. Color coding for the protein is as in Figure 1A. Secondary structure elements are labeled. DNA-contacting loops are labeled for monomer I. The disordered loop HI is drawn as a dashed line. (**B**) Stereo view of rearrangement of the N-terminal domains upon DNA binding, based on superposition of the C-terminal domains of both crystallographically independent monomers. The protein chains are drawn as Cα-traces (monomer I and II as thick and thin lines, respectively). The view is approximately perpendicular to that in (A). The DNA is shown in its relative position to monomer I. The α-helices A and B and the N- and C-termini are labeled. (**C**) Structural comparison of the N-terminal domains of NF-κB p52 (yellow) and NF-κB p50 (green) based on superposition of Cα atoms of the N-terminal domain core. The view is similar to that in (B). The protein structures are represented as backbone traces. The helices within the insert region are emphasized.

the NF-κB p50 complex. All secondary structure elements have been conserved between NF-κB p50 and p52, except for β-strands c and c′ in the C-terminal domain, which are both two residues shorter in p52. We used the p50 human homologue assignment of secondary structure elements. Major structural differences between p52 and p50 are only observed within the insert region (see below). The r.m.s. deviations for Cα atoms of corresponding residues in p50 and p52 are 0.9 Å for the core of the N-terminal domain (131 Cα atoms, insert region omitted), 1.0 Å for the C-terminal domain (101 Cα atoms) and 4.2 Å for the insert region (45 Cα atoms).

The asymmetric unit contains one homodimer–DNA complex. The two monomers are roughly related by a non-crystallographic dyad but they differ in the relative orientation of their N-terminal domains with respect to their C-terminal domains (Figure 2B). Taken individually, corresponding domains in the two monomers are virtually identical (r.m.s. deviation in Cα atom positions are 0.4 and 0.3 Å for the N-terminal and C-terminal domains,

respectively). Slight deviations from non-crystallographic symmetry (NCS) are observed in several loops (loops FG, G′αA, cc′ and fg). These loops are flexible and show high *B*-factors (see Materials and methods and Figure 7). Since the two monomers were refined individually, the mean r.m.s. deviation gives an estimate of the overall coordinate error. In rigid regions of the complex, namely the protein–DNA and the dimer interfaces, it is expected to be below 0.3 Å. Based on the superposition of the C-terminal domains, the reorientation of the N-terminal domain of monomer II, with respect to monomer I, involves a rotation about an axis which is perpendicular to the DNA helical axis and runs approximately through the interdomain linker. Additionally, the N-terminal domain of monomer II appears slightly shifted away from the DNA along the rotation axis. These domain movements are accommodated by a conformational change in the interdomain linker, namely at residues Ser226 and Asn227. By similar domain movements, accommodated by the linker, NF-κB p50 is able to recognize a half-site spacing of 3 bp (human homologue) and 4 bp (mouse homologue, compared in Müller *et al.*, 1996). The slight rearrangement of the domains in NF-κB p52 coincides with non-symmetrical bending of the DNA (see below).

Insert region

NF-κB p50 and p52 contain additional amino acid residues ('insert') which are located within an extension between β-strands G′ and H of the N-terminal domain, known as the 'insert region'. The insert region shows very little sequence similarity amongst different Rel family members and varies in length between 66 residues in NF-κB p50, 47 residues in NF-κB p52, 36 residues in Dorsal and 34 residues in NF-κB p65. The insert region of NF-κB p52 forms an autonomous module of two helices connected by a short loop (Figure 2). They pack tightly against each other and enclose an angle of about 50° (~40° in the structure of human p50). The two α-helices have been conserved between NF-κB p50 and p52 despite the low sequence similarity (only two residues in each helix are identical, Figure 1A). Helix αA points with its N-terminal end towards the phosphate backbone from the minor groove side.

The insert region of NF-κB p50 packs tightly against the outer concave surface of the N-terminal domain. In NF-κB p52, the insert region is rotated away from the core domain surface by about 20° opening up a deep, mainly polar cleft. The insert region module protrudes into the solvent and changes the shape of the complex considerably (Figure 2A, comparison in Figure 2C). Is the observed orientation of the insert region module an intrinsic structural feature of NF-κB p52? In human NF-κB p50, a total of 30 residues form the interface between the insert region and the N-terminal domain core (residues within 4.0 Å distance between the two elements). There are 12 hydrogen bonds formed by residues in helix αA (Arg157 and Arg164) and the connecting loop (Tyr166, Asn167, Leu171 and Gln180) with residues of the N-terminal domain core. In addition, several hydrophobic residues are buried in the interface (Ile97, Leu109, Ile123 and Leu171). In contrast, the p52 insert region element is connected to the domain core only by polar contacts of Lys153 in helix αA. This residue is either arginine or lysine in all Rel family members (Figure 1A). Furthermore, only 9 out of the 30 residues that pack the p50 insert region element onto the domain core are identical in the p52 sequence. However, since insert region elements of crystallographically related complexes pack against each other in both NCS-related monomers, we cannot rule out that their orientations are influenced by crystal packing forces. Nevertheless, the crystal contacts are not identical for both monomers, whereas the relative orientation of the insert region module with respect to the N-terminal domain core is exactly the same. This supports the argument that the different orientation of the insert region module in p52, as opposed to p50, is an intrinsic structural feature of the p52–DNA complex. Alternatively, it is still possible that the insert region is an autonomously moving element and that its orientation is ultimately defined through contacts with other interacting partners.

Interactions of the insert region with other DNA binding proteins

The insert region might interact with other DNA binding proteins. Due to the difference of the insert regions in NF-κB p50 and p52, the potential interaction surfaces presented to other macromolecules are substantially different. Interestingly, the largest movement of the insert region element of p52, relative to p50, is observed parallel to the DNA helical axis (Figure 2C). As a consequence, helix αB appears to be easily accessible to proteins adjacently bound to the DNA. Additionally, several mostly charged side chains (Gln176, Lys179, Glu180 and Lys183) point towards other factors potentially bound to a neighboring DNA binding site. The high-mobility group protein I(Y) [HMGI(Y)] contacts the minor groove at the AT-rich center of the PRDII κB site of the β-interferon promoter (Thanos and Maniatis, 1992). Residues of the insert region might be involved in HMGI(Y) interaction due to their proximity (Müller *et al.*, 1995). Furthermore, Rel proteins can interact with proteins of the basic region-leucine zipper (bZIP) families (Stein *et al.*, 1993a,b; Nolan, 1994). The monomeric transcription factor NF-ATc is related to the Rel family and the solution structure shows similarity of its DNA binding domain to the N-terminal domain of Rel proteins (Wolfe *et al.*, 1997). Site-directed mutagenesis experiments and model building lead to the conclusion that residues of NF-ATc corresponding to the Rel insert region are responsible for co-operative interaction with the bZIP transcription factor AP-1 on DNA in the interleukin-2 (IL-2) enhancer region (Chytil and Verdine, 1996; Peterson *et al.*, 1996; Wolfe *et al.*, 1997). The interleukin-8 (IL-8) promoter contains adjacent binding sites for Rel dimers and the bZIP transcription factor C/EBP (Stein and Baldwin, 1993). Compared with the IL-2 enhancer sequence, the sites are separated by three to four additional bp. The larger site separation is plausible since binding of a dimeric Rel protein, as opposed to monomeric NF-ATc, requires more space. Assuming a standard B-DNA conformation, the larger site separation would cause a relative rotation of the factors around the DNA helical axis by ~100–140°, as compared with the model for NF-ATc and AP-1 on the IL-2 enhancer (Peterson *et al.*, 1996). In such a model, the leucine zipper domain of C/EBP would be closest in space to the C-terminal domains of a Rel dimer. Indeed, a deletion of residues Glu222-Pro231 of p65 (correspond-

Table I. Interactions in the dimer interface

Residue in monomer I	Buried surface (Å^2)[a]	Conserved residue[b]	Hydrogen bonds to monomer II[c]		Van der Waals' contacts[e]
			Atom...residue atom	Distance (Å)[d]	
Ser231	14/4	–	–		His282 (1)
Arg232	101/12	◆	NE...Tyr247 OH	3.4 (3.4)	Tyr247 (2); Val288 (1)
			NH1...Asp280 OD1	3.3 (2.9)	
			NH2...Glu245 OE1	3.5 (3.6)	
Met233	11/2	–	N...Tyr247 OH	3.9 (3.4)	Tyr247 (4)
			O...Tyr247 OH	3.1 (3.2)	
Asp234	54/3	–	OD1...Asp234 OD1	2.6	Asp234 (1); Tyr247 (4)
			OD1...Asp234 OD2	3.0	
Glu245	22/0	□	see Arg232		–
Tyr247	110/64	–	see Arg232/Met233		Met233 (3); Asp234 (3); Tyr247 (8); Leu249 (6)
Leu249	73/73	◆	–		Tyr247 (4); Leu249 (1); His282 (12); Ala286 (1); Val288 (1)
Cys250	9/1	◆	N...His282 NE2	3.4 (3.4)	
			O...His282 NE2	2.8 (2.9)	His282 (2)
Asp251	32/0	□	OD1...Lys283 NZ	2.8 (-)	Lys283 (1)
Thr279	5/0	–	–		–
Asp280	20/1	◆	see Arg232		–
His282	92/56	◆	NE2...Tyr285 O and see Cys250	2.9 (3.1)	Ser231 (1); Leu249 (10); Cys250 (1); Tyr285 (3)
Lys283	60/53	–	see Asp251		Tyr285 (14); Asp251 (1)
Tyr285	81/73	–	see His282		His282 (4); Lys283 (16); Tyr285 (6)
Ala286	12/12	◆	–		Leu249 (2)
Val288	22/22	□	–		Arg232 (4); Leu249 (1)

[a]Total/hydrophobic accessible surface area of the given residue buried upon dimer formation. Calculated with AREAIMOL, DIFFAREA and RESAREA (CCP4, 1994) using the algorithm of Lee and Richards (1971) and a probe radius of 1.4 Å. Values for residues in monomer II are very similar.
[b]Residues which are conserved in all known family members are marked with (◆). The only exception is the *Drosophila* protein, Dif, that has a phenylalanine residue in place of His282. Residues that are conserved in human family members are marked with (□).
[c]Polar contacts with distances ⩽ 3.5 Å.
[d]The length of the H-bond is given. The length of the equivalent interaction from monomer II to monomer I is given in parenthesis.
[e]Residues in monomer II showing carbon–carbon contacts within 4.5 Å are given. The number of C–C atom pairs is given in parenthesis.

ing to loop bc and β-strand c in the C-terminal domain) abolishes interaction with C/EBP (Stein *et al.*, 1993b). In a similar system, the leucine zipper region of the Epstein–Barr virus bZIP transactivator BZLF1 has been shown to interact with the C-terminal region of p65 (Gutsch *et al.*, 1994). In this model, the N-terminal region of C/EBP would be close in space to the Rel insert region.

Dimer interface

A total of 1400 Å^2 of solvent accessible surface area is buried in the dimer interface. Only 50% of the dimer interface surface area is hydrophobic and many intermolecular hydrogen bonds are formed (Table I, Figure 3). Residues that only make van der Waals' contacts (Leu249, Ala286 and Val288), and most of the residues making important inter-dimer H-bonds (Arg232, Glu245, Cys250, Asp251, Asp280 and His282), are conserved amongst human Rel family members, while residues that contact the other monomer only through water molecules at the edge of the dimer interface (Ser231, Asp234 and Arg290) differ in different Rel proteins (Figure 1A, Table I). Residue Arg308 in human p50 makes extensive contacts to the DNA backbone. In p52, these interactions are lost and the corresponding residue Lys283 contacts the other monomer: In monomer II, the side chain NZ forms a hydrogen bond to the side chain of Asp251, and in monomer I it interacts via a tightly bound water molecule with the backbone amine group of Lys252 and with the side chain of Asp251 (Figures 3 and 6).

Why are the p50/p65 and p52/p65 heterodimers thermodynamically preferred over the p50 and p52 homodimers? Residue Phe310 in NF-κB p50, corresponding to Tyr285 in p52 and Val248 in p65, has been identified as a key residue which controls the specificity of dimerization (G.L.Verdine, unpublished data). This residue is located at the lower part of the dimer interface (Figure 3) and is not conserved among Rel family members. In NF-κB p52, Tyr285 packs tightly between the aliphatic side chain of Lys283 and the dimerically related tyrosine. This residue and Leu249 contribute most to the hydrophobic interface (Table I). In p50, the corresponding residue Phe310 adopts a very similar conformation. Another important residue, Tyr247, makes several polar interactions with its hydroxyl group (Table I). In p65, this residue is a phenylalanine, thus leading to deletion of the H-bonds in one half of a p52/p65 or a p50/p65 heterodimer. This seems contradictory to the observation that Rel heterodimers containing p65 are more stable than the corresponding homodimers. However, the asymmetries in a potential heterodimer interface may lead to conformational adjustments in the protein–DNA interface and thus to changes in site specificity.

DNA conformation

In the complex, the DNA is bound preferentially in one orientation as judged from the electron density for the

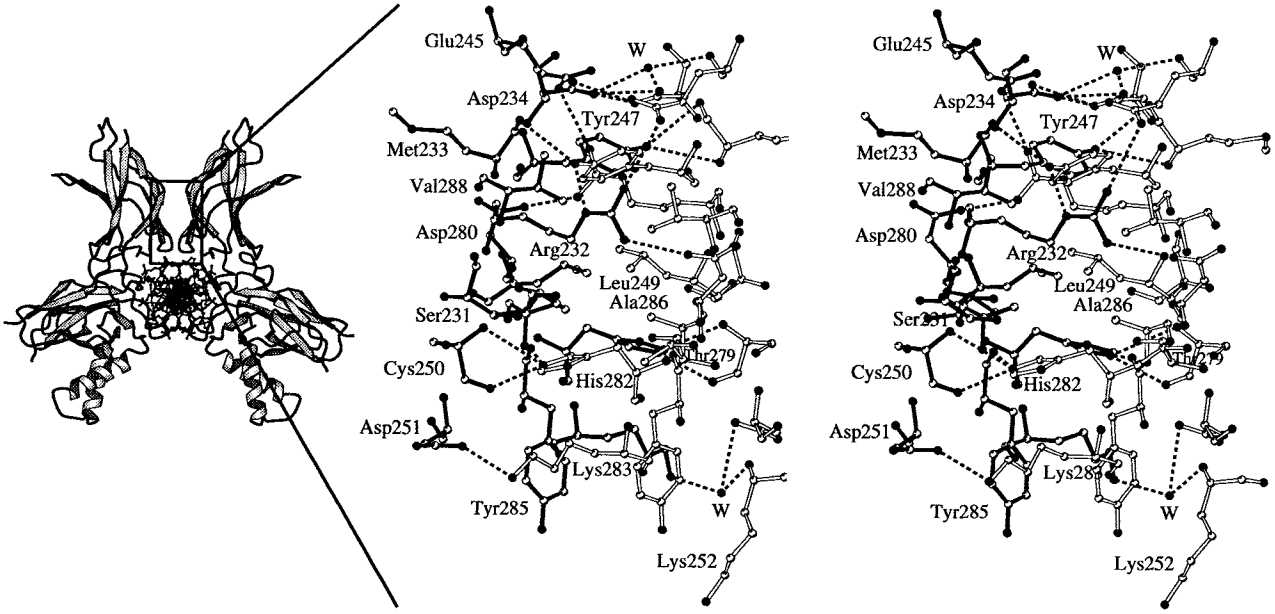

Fig. 3. Stereo view of residues in the dimer interface. Residues in monomer I and monomer II are drawn with filled and open bonds; carbon and 'polar' atoms as open and filled spheres, respectively. Intermolecular hydrogen bonds are drawn as dashed lines. Two bridging water molecules (W) are depicted as black spheres.

central base pair. A difference Fourier map calculated with phases from a model where these central bases are exchanged shows peaks indicating the wrong orientation. The preferred orientation is observed although the ends of the DNA are disordered and not involved in crystal contacts.

The r.m.s. deviation of all DNA atoms from canonical B-DNA is 1.4 Å after least squares fit of the phosphorous atoms. The overall geometry of the DNA fragment is similar to that in the human NF-κB p50 homodimer–DNA complex structure. The DNA fragment is slightly unwound (Figure 4) with an overall twist of 10.7 bp/turn (canonical B-DNA has 10.0 bp/turn). The central part of the DNA duplex (Thy/Ade-1 to Ade/Thy1) has a narrow minor groove that shows a single 'spine of hydration'. Hydration patterns like the minor groove 'spine' have also been observed in solution and thus are an integral part of the structure (Berman, 1991, 1994). The minor groove width is 3.6 Å at Thy/Ade0 (shortest distance between backbone O4′ atoms minus 2.8 Å, 5.7 Å for canonical B-DNA). Some unusual H-bonds are formed between the DNA bases. The central region of the duplex shows large propeller twists (up to -30.2° at Thy/Ade0, Figure 4) which allow for the formation of a stabilizing three centered H-bonding network (Blackburn and Gait, 1996) between the N6 amide of Ade-1, the O4 carbonyl of its Watson–Crick partner Thy-1 (3.2 Å) and the O4 carbonyl of the neighboring Thy0 (2.9 Å). Due to the lack of symmetry at the center of the DNA duplex, this three-centered interaction is unique to Ade-1 and impossible for the corresponding Ade-1 in the opposite strand. The large propeller twists of the central A:T bp are further propagated through base stacking over several bp (Figure 4). This leads to an unusual hydrogen bond between Gua-2 N2 and the O2 carbonyl of Cyt-3 (2.8 Å) and a corresponding interaction at the other half site (3.2 Å). The observed interactions may stabilize non-symmetric DNA distortions which could be recognized by indirect readout. Indeed, the level of propeller twist shows an inverse correlation with the flexibility of the dinucleotide step (Hassan and Calladine, 1996). The observed features of the central DNA sequence are typical for A-tracts (Dickerson *et al.*, 1994).

The NF-κB p52–DNA complex shows an overall DNA bend of 20° towards the C-terminal domain of monomer II (Figures 2C and 4). In comparison, an overall DNA bend of 30° (two 15° bends) towards the center of the C-terminal domains has been observed in the human NF-κB p50–DNA complex structure. The asymmetry of the bend in the p52 complex is probably caused by differences in the distortability of the pseudosymmetric DNA fragment resulting from the asymmetry of the central ATT sequence. In consequence, a slight adjustment of the recognition loop AB leads to slightly different orientations of the N-terminal domains at the two half sites. These adjustments most dramatically effect the position of the insert region which is most distant from the axis of rotation (Figure 2B). The maximum displacement between Cα atoms (5.8 Å) is observed for Arg160 at the C-terminal end of helix αA.

Details of DNA recognition

A total of 3200 Å^2 of accessible surface area is buried upon complex formation, accounting for the low dissociation constant of 15 pM for the MHC H-2 DNA site (Duckett *et al.*, 1993). Approximately 45% of the protein accessible surface area buried upon DNA binding is hydrophobic (Table II). A schematic diagram summarizing the observed protein–DNA interactions in both half sites is shown in Figure 5A and all observed protein–DNA contacts are listed in Table II. NF-κB p52 recognizes four concomitant guanines through side chain interactions of His62, Arg54, Arg52 and Lys221, respectively (Figure 5B). Presented by the 'recognition loop', the planar head groups of His62,

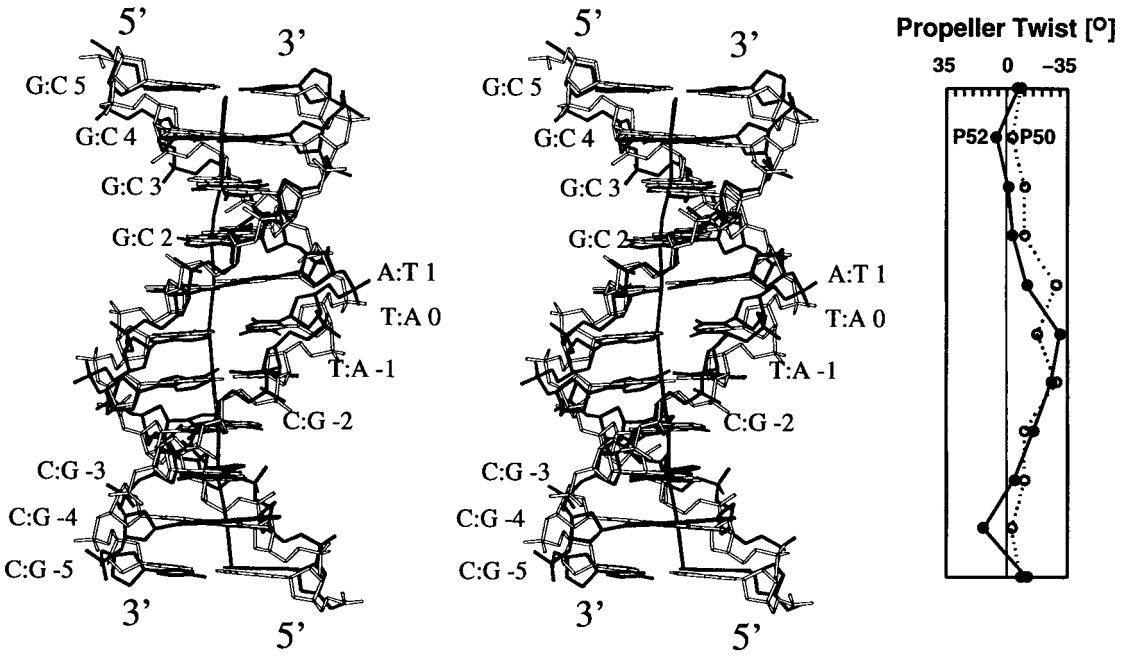

Fig. 4. Stereo representation of the 11 bp DNA duplex and comparison with canonical B-DNA. The DNA duplex as observed in the co-crystal structure is drawn with filled sticks. A corresponding sequence of canonical B-DNA (twist angle 36°, rise 3.38 Å) has been superimposed based on the phosphorous atoms and is drawn with open sticks. The view is onto the central minor groove (from below compared with Figure 2A). The propeller twist of each DNA base pair in the duplex of the NF-κB p52–DNA complex structure is plotted on the right (filled circles, solid line). For comparison, the propeller twist for corresponding bp in the human NF-κB p50–DNA complex structure is given (open circles, dashed line). The parameters for the human p50 homodimer–DNA complex structure show symmetry since the dimer dyad is crystallographic. Geometrical parameters and the helical axis were calculated with the program CURVES (Lavery and Sklenar, 1988).

Arg54 and Arg52 form a rigid stack. Additionally, Glu58 binds to the opposing cytosine Cyt ± 3 and buttresses the two arginines (Cyt ± 2 and Cyt ± 3 in human p50). In both half sites, the side chain of Cys57 hydrogen bonds to the backbone phosphate of Cyt2. This residue is suspected to be involved in redox-control of the DNA binding activity (Kumar *et al.*, 1992).

Most of the direct DNA contacts are conserved between NF-κB p52 and p50. However, Arg52 and Arg54 in NF-κB p52 show a slightly different pattern of interactions than in p50. In the NF-κB p50 structures, both corresponding arginines make bidentate interactions with N7 and O6 of two subsequent guanines through their guanidinium head groups. Compared with human p50, the guanidinium head group of Arg52 in p52 is rotated by ~15° around the NE–CZ bond. This allows the side chain of Arg52 to make polar contacts to three subsequent G:C bp in the major groove, at N7 of Gua4, O6 of Gua3 and N4 of Cyt2 (Figure 5, Table II). However, the contacts to Gua4 and Cyt2 are expected to be weaker due to the unfavorable H-bonding angle (Table II). Whereas Arg52 forms two favorable H-bonds with the O6 of Gua3, the N7 is not contacted. In the first half site, two structural water molecules help to orient the guanidinium head groups of Arg52 and Arg54 (Figure 5B). Water molecule 2 forms four H-bonds with favorable angles (Table II) and lies roughly in the planes of the three planar groups to which it binds (headgroups of Arg52 and Glu58 and the pyrimidine ring of Cyt2). Arg52 in monomer II shows slight differences in its H-bonding pattern (Table II) which could arise from differences in DNA conformation (see above). No equivalent water molecules are found in the second half site. There are some additional differences between p52 and p50. In the human NF-κB p50 homodimer–DNA complex structure, His144 and Arg308 make contacts to the DNA backbone. The corresponding residues His140 and Lys283 in p52 do not make any DNA interactions. Instead, the side chain of His140 hydrogen bonds to the backbone carbonyl of the neighboring residue Val141 while Lys283 forms a bridge between the two protein monomers (see above).

Most of the direct protein–DNA interactions are observed in both half sites. However, there are differences which probably result from the relative reorientation of the N-terminal domains. In the second half site, the interaction of His62 with the outermost guanidine is lost but additional contacts to the DNA backbone are observed (Table II). A set of van der Waals' interactions is observed in both half sites (Table II). The importance of the packing of Pro223 against the ribose of Ade1 and of the aromatic ring of Tyr55 against Thy1 as positioning contacts has been noted (Müller *et al.*, 1995). Most of the residues contacting DNA are well buttressed. For example, residues Val138 to His140 make multiple interactions with the recognition loop and establish a connection to helix αA of the insert region. Many water molecules participate in this buttressing network.

Role of water in DNA binding

A total of 20 well-ordered water molecules with an average *B*-factor of 41.5 Å^2 are located in the protein–DNA interface (water molecules within 3.5 Å of both protein and DNA). On average, each water molecule forms three hydrogen bonds. Of these 20 water molecules, 14 make interactions with at least one polar protein and one polar DNA atom (Table II, Figure 5A). Among these 14 water

Table II. Protein–DNA contacts

DNA[a]	Protein	Distance[b] (Å) [angle[c] (°)]		Buried surface area (Å^2)[d]	
		Monomer I	Monomer II	Monomer I	Monomer II
A1. Polar interactions with DNA bases					
Gua5 O6	His62 ND1	3.4 (93)	4.4[e]		
Gua5 N7	Arg54 NH2	3.4 (114)	3.6 (120)	33/0	35/0
	His62 ND1	2.9 (123)	3.8[e]	65/32	50/19
Gua4 O6	Arg54 NH1	3.2 (157)	3.0 (160)		
Gua4 N7	Arg52 NH1	3.2 (96)	2.8 (110)	62/1	52/1
	Arg54 NH2	3.1 (148)	3.0 (145)		
Gua3 O6	Arg52 NH1	2.9 (144)	3.8 (136)		
	Arg52 NH2	2.9 (146)	3.3 (159)		
Cyt3 N4	Arg52 NH2	3.4 (120)	3.8 (133)		
	Glu58 OE2	3.0 (149)	2.6 (148)	41/19	36/14
Gua2 O6	Lys221 NZ	2.9	2.7	64/21	80/41
Cyt2 N4	Arg52 NH2	2.9 (118)	3.1 (116)		
A2. Polar interactions with the DNA backbone					
Cyt2 O1P	Cys57 SG	3.4 3.5	67/22	70/21	
Cyt2 O5′	Cys57 SG	3.4	3.3		
Thy1 O1P	Tyr55 OH	2.7 (170)	2.7 (170)	78/44	74/39
Thy1 O2P	Lys143 N	2.9 (174)	3.0 (178)		
Thy1 O5′	Tyr55 OH	3.4 (90)	3.6 (90)		
Ade/Thy0 O1P	Lys143 NZ	4.6	3.5	82/35	66/28
Ade-1 O1P	Gln284 NE2	3.2 (165)	3.2 (156)	44/8	61/22
Ade-1 O2P	Gln254 NE2	4.3 (128)	2.8 (133)	8/4	21/5
	Gln284 NE2	3.5 (133)	3.5 (115)		
B. Water-mediated interactions[f]					
Cyt3 O1P	W1-Cys57 O	2.8–2.6	–		
Cyt2 N4	W2-Arg52 NH2	3.0–3.2	–		
	W2-Arg52 NE	3.0–3.1	–		
	W2-Glu58 OE1	3.0–3.0	–		
Cyt2 O1P	W3-Gly56 N	2.9–3.1	–		
	W3-Cys57 N	2.9–2.9	–		
Thy1 O2P	W4-Lys143 NZ	3.3–3.3	–		
Thy1 O2P	W5-Lys143 NZ	3.4–2.6	–		
Ade0 O1P	W6-Ser222 OG	2.8–2.6	–		
Ade0 O3′	W4-Lys143 NZ	3.3–3.3	–		
Thy0[g] O1P	W7-Lys143 NZ	–	3.0–2.7		
Thy0[g] O4	W8-Lys252 NZ	–	2.9–3.4		
Ade1[g] O1P	W9/W9′-Lys252 NZ	2.7–2.7	2.7–2.7		
Gua2[g] O2P	W10-Val281 O	–	2.6–3.3		
	W10-Gln284 N	–	2.6–2.6		
	W11-Lys255 N	–	2.8–2.8		
Gua4[g] O1P	W12-Gly50 O	–	3.1–3.5		
	W12-Asn227 ND2	–	3.1–3.5		
	W13-Ser220 OG	–	3.5–3.3		
C. Van der Waals' interactions[h]					
Gua5	Gly63			10/10	7/7
Gua4	Gly50			12/0	6/0
Thy1	Thr142			32/32	33/33
	Lys144			25/25	47/23
Ade-1	Pro223			59/59	53/53
Gua-2	Lys252			5/0	23/0
	Lys255			24/12	43/24
Gua-3	Lys283			59/19	54/24

[a]Nucleotides listed are contacted by monomer I unless stated otherwise. The corresponding interaction with monomer II is also given.
[b]Polar contacts with distances ⩽ 3.5 Å in at least one half site.
[c]The hydrogen bonding angle is defined as donor-H...acceptor.
[d]Total/hydrophobic accessible surface area of the given protein residue buried upon DNA binding. Calculated with the programs AREAIMOL, DIFFAREA and RESAREA (CCP4, 1994) using the algorithm of Lee and Richards (1971) and a probe radius of 1.4 Å. The values are given only when a residue is listed for the first time.
[e]Compared with the first half site, the side chain of His62 is rotated by 180° in the second half site to allow for H-bonding interactions of ND1 to the protein backbone. For comparison, the distance from Gua5 N7 to His62 CE2 is given.
[f]The numbering of water molecules corresponds to that used in Figure 5A.
[g]Nucleotides contacted by monomer II.
[h]Listed are residues which are only involved in van der Waals' interactions. Only protein residues which show at least 10 Å^2 of total surface area buried in at least one half site are listed.

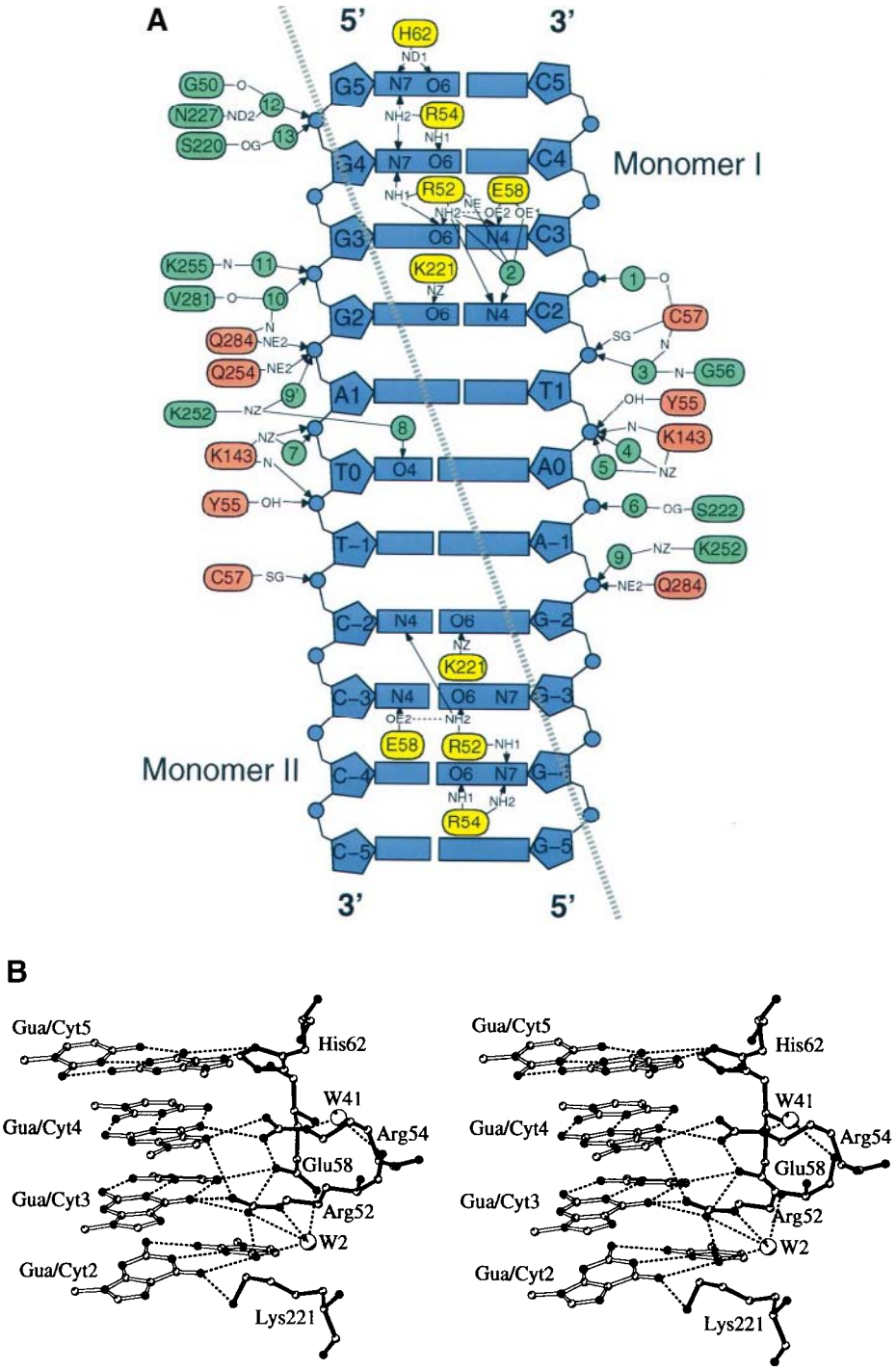

Fig. 5. DNA recognition by the NF-κB p52 homodimer. (**A**) Schematic diagram of polar interactions between protein and DNA. Contacts to DNA bases (yellow boxes), to the DNA backbone (red boxes) and water-mediated contacts (green boxes) in both half sites are shown. Water molecules are depicted as green spheres. (**B**) Stereo view of DNA base-specific recognition in the first half site. Protein residues and DNA bases are drawn with filled and open bonds, respectively. Water molecules are drawn as open spheres. DNA backbone atoms have been omitted for clarity. Polar interactions are indicated as broken lines.

molecules, only one pair of NCS-related waters is found (9/9′, Table II). Thus, water-mediated interactions are substantially different in the two half sites. The remaining six water molecules contact either polar protein or DNA atoms and make stabilizing interactions to neighboring waters in the protein–DNA interface. Almost all water-mediated contacts are to the sugar–phosphate backbone (Figure 5A, Table II). Exceptions are found at Arg52 in the first half site (see above) and at Lys252 in the second half site which contacts the O4 carbonyl of Thy0 via water molecule 8 in the major groove of the central core region (Figure 5A). The interaction of Lys252 to the phosphate of Ade-1 is mediated by a tightly bound water molecule in both half sites in our structure while it appeared to be direct in the human NF-κB p50 homodimer–DNA complex. Residues Ser220, Ser222 and Asn227 in

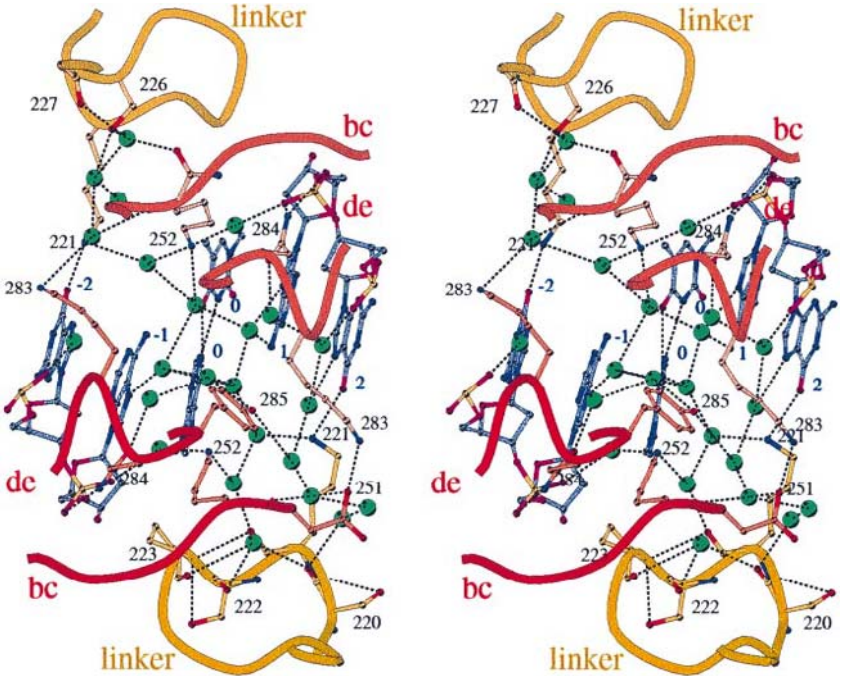

Fig. 6. Stereo view of the solvent network in the interfacial water cavity between the major groove of the central DNA sequence and the edge of the dimer interface. The view is along the approximate dimer dyad (from above as compared with Figure 2A). Loops from monomer I and II are show on the bottom and the top, respectively. Twenty four ordered water molecules (green spheres) within a radius of 8 Å from the hydroxyl groups of Tyr285 residues are shown. Hydrogen bonds are indicated as dashed lines. Only protein residues which participate in the H-bonding network are shown. DNA nucleotides that are not contacted have been omitted for clarity.

the interdomain linker make water-mediated contacts to the DNA backbone. These residues are changed to asparagine, alanine and glutamate in p65 which might cause differences in positioning of p52/p65 or p50/p65 heterodimers.

A large interfacial water cavity is located above the major groove in the central part of the DNA (Figure 6). A total of 24 water molecules establish a complicated network of interwoven hydrogen bonds between the edge of the dimer interface (loops bc and de), the interdomain linker and DNA bp Gua/Cyt-2 to Cyt/Gua-2 (Figure 6). The mean *B*-factor for these water molecules is 40.5 Å^2 and thus substantially lower than the average for all water molecules included in the model (48.9 Å^2). This solvent network shows little symmetry. Although superposition of NCS-related residues in loops bc and de puts six water molecules in a distance less than 1.5 Å to another water molecule, only the already mentioned water molecules 9 and 9′ (Figures 5A and 6) have the same H-bonding partners and can therefore be regarded as strictly NCS-related water molecules.

It is becoming well accepted that single ordered water molecules bridging between polar protein and DNA atoms can contribute to the specificity of protein–DNA complexes (Schwabe, 1997). In the case of Trp repressor, detailed mutagenesis studies showed such a contribution (Bass *et al.*, 1988; Joachimiak *et al.*, 1994). Similarly, water-mediated contacts play an important role in DNA recognition by the paired homeodomain (Wilson *et al.*, 1995). The question arises whether polar protein atoms separated from DNA by two or three fixed water molecules can still contribute to site specificity. DNA has a sequence-dependent hydration pattern which in the proximity of polar protein atoms might give rise to preferred water networks that could contribute energetically to site recognition. For example, Tyr285 in monomer I forms H-bonds, with good stereochemistry, to two waters (Figure 6). These waters interact with several other waters, amongst them two waters which are hydrogen-bonded to Gua-2 N7 and Ade0 N6, respectively. Recently, the energetic contribution of an interfacial water cavity to homeodomain DNA binding has been examined (Labeots and Weiss, 1997). Based on stereoselective chemical modification of phos-

Table III. Statistics of data processing and refinement

Data collection	
Space group	$P2_12_12_1$
Unit cell dimensions (Å)	44.2×121.0×134.9
Resolution (Å) (highest resolution shell)	20.0–2.1 (2.2–2.1)
Number of measured reflections	124281 (8537)
Number of unique observations	39642 (4127)
Redundancy	3.1 (2.1)
Completeness (%)	91.8 (78.1)
R_{sym} (%)[a]	6.1 (31.2)
Refinement	
Total number of protein residues	568
Number of nucleotides	22
Number of solvent molecules	785
Total number of non-hydrogen atoms	5736
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.013
Bond angles (°)	2.1
Average overall isotropic *B*-factor (Å^2)	
—protein (main chain/side chain)	45.6 (46.7/45.1)
—DNA	43.5
—Solvent molecules	48.9
R_{cryst}/R_{free} (10.0–2.1 Å) [%]	21.9/32.0

[a]$R_{sym} = \Sigma|I_i-\langle I_i\rangle|/\Sigma|\langle I_i\rangle|$, where I_i is the intensity of the individual reflection and $\langle I_i\rangle$ is the mean value of its equivalent reflections.
Values given in parenthesis correspond to the highest resolution shells.

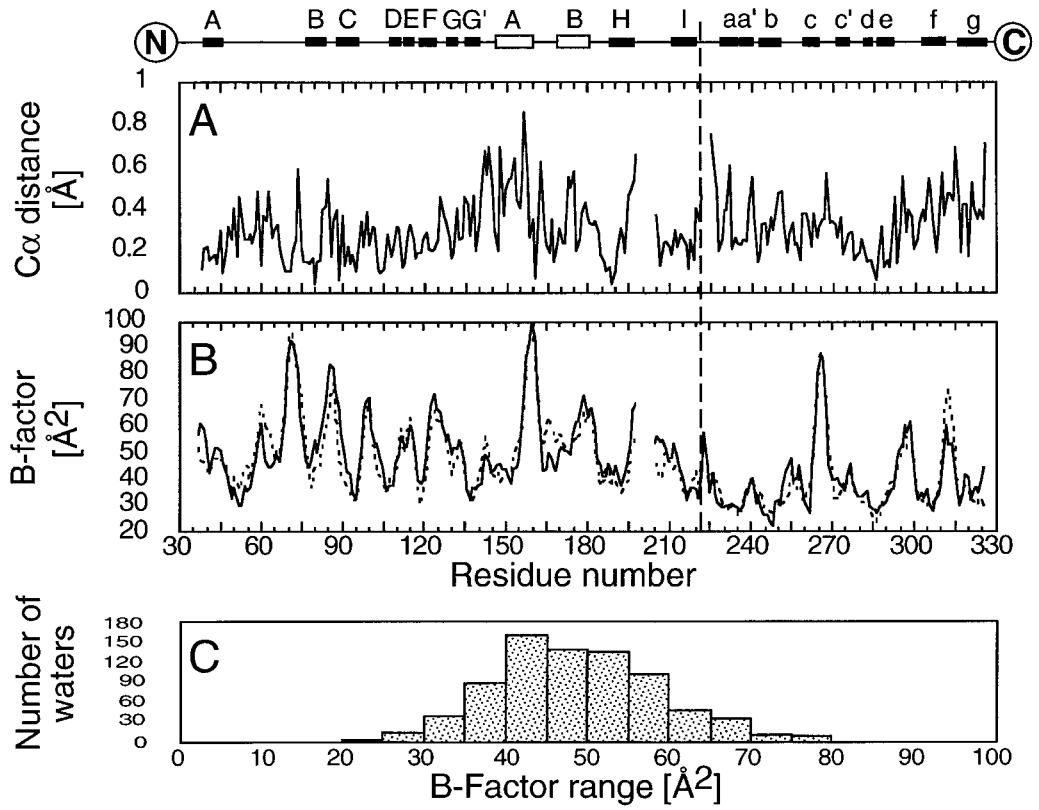

Fig. 7. (**A**) Distance of Cα atoms in the two monomers after separate superposition of the N-terminal and the C-terminal domains. Secondary structure elements are indicated above the plot. Black bars indicate β-strands, white boxes indicate α-helices. The dashed line marks the border between the N-terminal and the C-terminal domain. (**B**) Thermal mobility of the polypeptide chains. Shown are the average *B*-factors of main chain atoms in monomer I (solid line) and monomer II (dashed line). The overall *B*-factor for the N-terminal domains is higher than that for the C-terminal domains. (**C**) *B*-factor distribution of the 785 water molecules included in the model.

phate groups, the reported results provide evidence for the long-range importance of an ordered water structure in the interfacial cavity.

Site specificity

The homodimeric complex of NF-κB p52 shows some asymmetry, which might ultimately result from the central asymmetry of the DNA binding site. Unsymmetrical DNA bending, asymmetric positioning of the N-terminal domains, slightly different protein–DNA interactions at the two half sites and an asymmetric water structure are observed. The asymmetry extends into the cavity above the central ATT sequence. Ordered water molecules occupy this cavity between the edge of the dimer interface and the major groove of the DNA. Chang *et al*. (1994) suggested that bases within the central part of the binding site influence the DNA binding affinity of NF-κB p52, although these bases are not directly contacted by the protein. Furthermore, it was shown that the difference in binding specificity of p50 and p52 is located in the C-terminal 50 amino acid residues of the RHR. Swapping part of the C-terminal domain of p50 (residues 298–434) to p52 increased the affinity of the chimeric molecule to the Ig/HIV site to the level of p50 (Schmid *et al.*, 1994). The swapped region contains loop de which helps to establish the water network described above. In this loop, Tyr285 in NF-κB p52 (Phe310 in p50, Val248 in p65) appears to play a key role. Interestingly, this residue also forms part of the dimer interface and controls dimerization specificity (see above). Its two functions in dimerization and in establishing a water network strengthen the view of the dimer and protein–DNA interfaces as one continuous recognition surface (Müller *et al.*, 1995). We predict that the interfacial water cavity connecting the C-terminal domains and the central DNA bases might be important for the discrimination of different sites and that the difference in specificity between NF-κB p52 and p50 is largely caused by Tyr285. Further biochemical experiments as site-directed mutagenesis of residues involved in the formation of the water network, in particular of Tyr285, as well as chemical probing of the contacted DNA as described (Labeots and Weiss, 1997), have to prove this hypothesis.

Materials and methods

Crystal structure solution

Bacterial expression and purification of human NF-κB p52 (residues 35–329) containing the entire RHR apart from the C-terminal nuclear localization signal and co-crystallization with a 13mer DNA fragment comprising the natural κB binding site MHC H-2 (Figure 1B) have been described in detail (Cramer and Müller, 1997). Diffraction data were collected at BM14 at the ESRF in Grenoble from a single crystal at 100 K using a MAR image plate. Data processing was carried out with programs DENZO and SCALEPACK (Otwinowski and Minor, 1997). Crystals belong to space group $P2_12_12_1$ with unit cell dimensions 44.2×121.0× 134.9 Å. The asymmetric unit is occupied by one NF-κB p52 homodimer–DNA complex and the solvent content is 48.7% (V_M = 2.40 Å^3/Da; Matthews, 1968). The obtained dataset to 2.1 Å resolution is 91.8% complete with R_{sym} = 6.1% on intensities (Table III). The

structure was determined by molecular replacement using the program suite AMoRe (Navaza, 1994). Rotation and translation function searches within various resolution ranges (10.0–5.0, 10.0–4.5 and 10.0–4.0 Å) were carried out with several different search models based on the human NF-κB p50 homodimer–DNA complex structure (Müller *et al.*, 1995). A dimeric search model including the DNA in which the entire insert region (residues 145–210) had been omitted and all differing side chains had been truncated to alanine gave a unique solution which could be selected amongst 50 possible solutions based on the *R*-factor, the correlation coefficient and reasonable packing interactions. Using reflections from 10.0 to 4.0 Å resolution, an integration sphere radius of 30 Å and sampling steps of 2.5°, an *R*-factor of 52.9% and a correlation coefficient of 0.33 were obtained for this solution after the translation search. After rigid body refinement with AMoRe, the *R*-factor decreased to 45.9% and the correlation coefficient increased to 0.48 (for comparison, an *R*-factor of 53.9% and a correlation coefficient of 0.24 were obtained for the next highest peak). Visual inspection of the molecular replacement solution for the dimeric complex and the symmetry mates using O (Jones *et al.*, 1991) revealed no bad contacts.

Structure refinement

Refinement was carried out using the programs X-PLOR (Brünger, 1992) and O (Jones *et al.*, 1991) as follows. 10% of the measured reflections were excluded from the refinement process and used for cross-validation with the free *R*-factor. For refinement of the DNA with X-PLOR, improved stereochemical parameters were used (Parkinson *et al.*, 1996). The first model contained 11 bp of DNA and lacked 47 residues of the insert region (residues 141–187) and 89 side chains per monomer. All individual *B*-factors were set to 30 Å^2. Rigid body refinement of five independent groups (two N-terminal domains, two C-terminal domains and the DNA duplex) was carried out using reflections from 10.0 to 4.5 Å and gave a free *R*-factor of 44.5% (R = 41.8%). In the following refinement cycles, non-crystallographic symmetry restraints (weight = 300 kcal mol^{-1} Å^{-2}) were applied separately to the N- and C-terminal domains. Positional refinement using data from 10.0–3.0 Å gave a free *R*-factor of 44.1% (R = 36.9%). For the resulting model, the overall *B*-factor followed by individual isotropic *B*-factors were refined using restraints of 1.5 Å^2 for main chain and 2.5 Å^2 for side chain atoms. Subsequent positional refinement yielded a free *R*-factor of 43.4% (*R* = 30.3%). After solvent correction, electron density maps with Fourier coefficients ($2F_o$–F_c) and (F_o–F_c) were calculated using phases from the obtained model. These first maps showed clear electron density for 38 missing side chains per monomer and for about half of the residues belonging to the insert region. After several rounds of model building followed by positional and individual *B*-factor refinement, the resolution range of the data was extended to 10.0–2.5 Å. The quality of the electron density maps improved further so that all missing side chains and the entire insert region could be built in. The NCS restraints were loosened (weight = 20 C) and several loop regions showing deviations from NCS were rebuilt. The resolution range was subsequently further extended to 10.0–2.1 Å (39 039 unique reflections). Possible water molecules were located in difference Fourier maps with the programs PEAKMAX and WATPEAK (CCP4, 1994). Peaks of a height greater than 3σ were inspected graphically and water molecules with an initial *B*-factor of 30 Å^2 were included in the model if reasonable H-bonds were formed (2.5–3.5 Å distance from polar atoms of the protein, the DNA or another molecule). Water molecules were deleted when their *B*-factor rose over 60 Å^2 after refinement. In later stages of refinement, water molecules with *B*-factors of 60–80 Å^2 after refinement were kept when they showed good H-bonding geometry and substantial ($2F_o$–F_c) electron density over 1.0σ. For the last rounds of refinement the restraints for individual *B*-factor refinement were loosened (2.5 and 3.5 Å^2 for main chain and side chain atoms, respectively). The data were corrected for weaker diffraction along the *a*-axis (the direction of the helical axis of the DNA) by refinement of an overall anisotropic *B*-factor using X-PLOR. This improved the free *R*-factor by 0.5%. Additionally, local scaling of F_{obs} to F_{calc} was carried out with MAXSCALE (M. Rould, personal communication) to correct for anisotropic diffraction. This correction lowered the free *R*-factor by 0.8%. No NCS restraints were applied in the last refinement cycles. Four terminal nucleotides outside the κB binding site (Figure 1B), residues 200–205 (loop HI) as well as two residues at each terminus are disordered in both monomers and were excluded from the final model (Figure 1A). The refined model contains 284 residues per protein monomer, 11 bp of DNA and 785 water molecules and has a crystallographic *R*-factor of 21.9% (free *R*-factor = 32.0%). The coordinates and structure factors will be deposited with the Brookhaven Protein Data Bank.

Structure analysis

The quality of the final model was assessed using programs PROCHECK (Laskowski *et al.*, 1993) and WHATIF (Vriend, 1990). Both protein chains in the asymmetric unit show good stereochemistry with 85.6% of the non-glycine and non-proline residues in the most favored regions of the Ramachandran plot. Only Arg160 in monomer I and Asn145 and Gln284 in monomer II lie in the generously allowed regions. Arg160 at the outermost end of helix αA in monomer I is part of a badly ordered region; its side chain is disordered (compare Figure 7B). For Asn145 and Gln284 in monomer II there is clear ($2F_o$–F_c) electron density which shows that they are correctly modeled. The r.m.s. deviation from ideal geometry is 0.013 Å for bond lengths and 2.1° for bond angles (Table III). The mean *B*-factor is 46.7 Å^2 for main chain and 45.1 Å^2 for side chain protein atoms with only minor differences between the two monomers (Table III). The distribution of *B*-factors along the protein chains is shown in Figure 7B. Interestingly, the C-terminal domains show lower overall *B*-factors than the N-terminal domains (38.6 Å^2 versus 51.4 Å^2 for main chain atoms). While the C-terminal domains are tightly packed against each other and make extensive crystal contacts, the N-terminal domains appear to be more flexible. Whereas protein residues close to the DNA axis and in and around the dimer interface are highly ordered, residues 69–71 of loop AB, loops BC, CD, HI and cc′, and the C-terminal end of helix αA, are badly ordered (Figure 7B). The flexibility of these loops might explain the relatively high free *R*-factor. Of the 785 water molecules included in the model, 61% belong to the first solvation shell (within 3.5 Å from protein or DNA), and 90% are within 5.0 Å of either protein or DNA. The *B*-factor distribution of the water molecules is given in Figure 7C. The average *B*-factor for all water molecules is 48.9 Å^2 which agrees with the overall *B*-factor for all protein and all DNA atoms (Table III). The geometry of the DNA was analyzed with CURVES (Lavery and Sklenar, 1988). Least squares fitting and calculation of r.m.s. deviations were carried out with LSQKAB (CCP4, 1994). Atomic distances and H-bonding angles were calculated with CONTACT (CCP4, 1994). Figures were drawn with Molscript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).

Acknowledgements

We would like to thank A.Thompson (EMBL Grenoble) for help at BM14 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. This research was supported by a John Dickson Fisher Research Grant from the American Foundation for AIDS Research (to G.L.V.).

References

Baeuerle,P.A. and Baltimore,D. (1996) NF-κB: ten years after. *Cell*, **87**, 13–20.

Baeuerle,P.A. and Henkel,T. (1994) Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.*, **12**, 141–179.

Baldwin,A. (1996) The NF-κB and IκB proteins: new discoveries and insights. *Annu. Rev. Immunol.*, **14**, 649–681.

Baldwin,A.,Jr and Sharp,P.A. (1987) Binding of a nuclear factor to a regulatory sequence in the promoter of the mouse H-2 κB class I major histocompatibility gene. *Mol. Cell. Biol.*, **7**, 305–313.

Baldwin,A.,Jr and Sharp,P.A. (1988) Two transcription factors, NF-κB and H2TF1, interact with a single regulatory sequence in the class I major histocompatibility complex promoter. *Proc. Natl Acad. Sci. USA*, **85**, 723–727.

Barillas-Mury,C., Charlesworth,A., Gross,I., Richman,A., Hoffmann,J.A. and Kafatos,F.C. (1996) Immune factor Gambif1, a new rel family member from the human malaria vector, *Anopheles gambiae*. *EMBO J.*, **15**, 4691–4701.

Bass,S., Sorrells,V. and Youderian,P. (1988) Mutant Trp repressors with new DNA binding specificities. *Science*, **242**, 240–245.

Berman,H.M. (1991) Hydration of DNA. *Curr. Opin. Struct. Biol.*, **1**, 423–427.

Berman,H.M. (1994) Hydration of DNA: take 2. *Curr. Opin. Struct. Biol.*, **4**, 345–350.

Blackburn,G.M. and Gait,M.J. (1996) *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford, UK.

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

Brünger,A.T. (1992) *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
CCP4 (1994) Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–776.
Chang,C.C., Zhang,J., Lombardi,L., Neri,A. and Dalla-Favera,R. (1994) Mechanism of expression and role in transcriptional control of the proto-oncogene NFKB-2/LYT-10. *Oncogene*, **9**, 923–933.
Chytil,M. and Verdine,G.L. (1996) The Rel family of eukaryotic transcription factors. *Curr. Opin. Struct. Biol.*, **6**, 91–100.
Cramer,P. and Müller,C.W. (1997) Engineering of diffraction-quality crystals of the NF-κB p52 homodimer–DNA complex. *FEBS Lett.*, **405**, 373–377.
Dickerson,R.E., Goodsell,D.S. and Neidle,S. (1994) '...the tyranny of the lattice...'. *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.
Duckett,C.S., Perkins,N.D., Kowalik,T.F., Schmid,R.M., Huang,E.S., Baldwin,A.,Jr and Nabel,G.J. (1993) Dimerization of NF-κB2 with RelA(p65) regulates DNA binding, transcriptional activation, and inhibition by an IκB-α (MAD- 3). *Mol. Cell. Biol.*, **13**, 1315–1322.
Franzoso,G., Bours,V., Azarenko,V., Park,S., Tomita-Yamaguchi,M., Kanno,T., Brown,K. and Siebenlist,U. (1993) The oncoprotein Bcl-3 can facilitate NF-κB-mediated transactivation by removing inhibiting p50 homodimers from select κB sites. *EMBO J.*, **12**, 3893–3901.
Ghosh,G., van Duyne,G., Ghosh,S. and Sigler,P.B. (1995) Structure of NF-κB p50 homodimer bound to a κB site. *Nature*, **373**, 303–310.
Gilmore,T.D. and Morin,P.J. (1993) The IκB proteins: members of a multifunctional family. *Trends Genet.*, **9**, 427–433.
Gutsch,D.E., Holley-Guthrie,E.A., Zhang,Q., Stein,B., Blanar,M.A., Baldwin,A.S. and Kenney,S.C. (1994) The bZIP transactivator of Epstein-Barr virus, BZLF1, functionally and physically interacts with the p65 subunit of NF-κB. *Mol. Cell. Biol.*, **14**, 1939–1948.
Hassan,M.A.E. and Calladine,C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*, **259**, 95–103.
Joachimiak,A., Haran,T.E. and Sigler,P.B. (1994) Mutagenesis supports water mediated recognition in the trp repressor-operator system. *EMBO J.*, **13**, 367–372.
Jones,A.T., Zhou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
Kumar,S., Rabson,A.B. and Gelinas,C. (1992) The RxxRxRxxC motif conserved in all Rel/κB proteins is essential for the DNA binding activity and redox regulation of the v-Rel oncoprotein. *Mol. Cell. Biol.*, **12**, 3094–3106.
Labeots,L.A. and Weiss,M.A. (1997) Electrostatics and hydration at the homeodomain–DNA interface: chemical probes of an interfacial water cavity. *J. Mol. Biol.*, **269**, 113–128.
Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.*, **6**, 63–91.
Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
Lin,R., Gewert,D. and Hiscott,J. (1995) Differential transcriptional activation *in vitro* by NF-κB/Rel proteins. *J. Biol. Chem.*, **270**, 3123–3131.
Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
Merritt,E.A. and Murphy,M.E.P. (1994) Raster3D Version 2.0-a program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.
Müller,C.W., Rey,F.A., Sodeoka,M., Verdine,G.L. and Harrison,S.C. (1995) Structure of the NF-κB p50 homodimer bound to DNA. *Nature*, **373**, 311–317.
Müller,C.W., Rey,F.A. and Harrison,S.C. (1996) Comparison of two different DNA-binding modes of the NF-κB p50 homodimer. *Nature Struct. Biol.*, **3**, 224–227.
Navaza,J. (1994) AMoRe: an automated program package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.
Nolan,G.P. (1994) NF-AT-AP-1 and Rel-bZIP: hybrid vigor and binding under the influence. *Cell*, **77**, 795–798.
Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
Parkinson,G., Vojtechovsky,J., Clowney,L., Brünger,A.T. and Berman, H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr.*, **D52**, 57–64.
Perkins,N.D., Schmid,R.M., Duckett,C.S., Leung,K., Rice,N.R. and Nabel,G.J. (1992) Distinct combinations of NF-κB subunits determine the specificity of transcriptional activation. *Proc. Natl Acad. Sci. USA*, **89**, 1529–1533.
Peterson,B.R., Sun,L.J. and Verdine,G.L. (1996) A critical arginine residue mediates cooperativity in the contact interface between transcription factors NFAT and AP-1. *Proc. Natl Acad. Sci. USA*, **93**, 13671–13676.
Rost,B. and Sander,C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.
Rost,B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.*, **266**, 525–539.
Schmid,R.M., Perkins,N.D., Duckett,C.S., Andrews,P.C. and Nabel,G.J. (1991) Cloning of an NF-κB subunit which stimulates HIV transcription in synergy with p65. *Nature*, **352**, 733–736.
Schmid,R.M., Liptay,S., Betts,J.C. and Nabel,G.J. (1994) Structural and functional analysis of NF-κB. Determinants of DNA binding specificity and protein interaction. *J. Biol. Chem.*, **269**, 32162–32167.
Schwabe,J.W. (1997) The role of water in protein–DNA interactions. *Curr. Opin. Struct. Biol.*, **7**, 126–134.
Stein,B. and Baldwin,A.,Jr (1993) Distinct mechanisms for regulation of the interleukin-8 gene involve synergism and cooperativity between C/EBP and NF-κB. *Mol. Cell. Biol.*, **13**, 7191–7198.
Stein,B., Baldwin,A.,Jr, Ballard,D.W., Greene,W.C., Angel,P. and Herrlich,P. (1993a) Cross-coupling of the NF-κB p65 and Fos/Jun transcription factors produces potentiated biological function. *EMBO J.*, **12**, 3879–3891.
Stein,B., Cogswell,P.C. and Baldwin,A.,Jr (1993b) Functional and physical associations between NF-κB and C/EBP family members: a Rel domain-bZIP interaction. *Mol. Cell. Biol.*, **13**, 3964–3974.
Thanos,D. and 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.
Thanos,D. and Maniatis,T. (1995) NF-κB: a lesson in family values. *Cell*, **80**, 529–532.
Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56.
Wilson,D.S., Guenther,B., Desplan,C. and Kuriyan,J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.
Wolfe,S.A., Zhou,P., Dotsch,V., Chen,L., You,A., Ho,S.N., Crabtree,G.R., Wagner,G. and Verdine,G.L. (1997) Unusual Rel-like architecture in the DNA binding domain of the transcription factor NFATc. *Nature*, **385**, 172–176.

Received on August 20, 1997.