Structure of the specificity domain of the Dorsal homologue Gambif1 bound to DNA

Patrick Cramer[1†], Annabelle Varrot[1‡], Carolina Barillas-Mury[2§], Fotis C Kafatos[2] and Christoph W Müller[1*]

Background: NF-κB/Rel transcription factors play important roles in immunity and development in mammals and insects. Their activity is regulated by their cellular localization, homo- and heterodimerization and association with other factors on their target gene promoters. Gambif1 from *Anopheles gambiae* is a member of the Rel family and a close homologue of the morphogen Dorsal, which establishes dorsoventral polarity in the *Drosophila* embryo.

Results: We present the crystal structure of the N-terminal specificity domain of Gambif1 bound to DNA. This first structure of an insect Rel protein–DNA complex shows that Gambif1 binds a GGG half-site element using a stack of three arginine sidechains. Differences in affinity to Dorsal binding sites in target gene promoters are predicted to arise from base changes in these GGG elements. An arginine that is conserved in class II Rel proteins (members of which contain a transcription activation domain) contacts the outermost guanines of the DNA site. This previously unseen specific contact contributes strongly to the DNA-binding affinity and might be responsible for differences in specificity between Rel proteins of class I and II.

Conclusions: The Gambif1–DNA complex structure illustrates how differences in Dorsal affinity to binding sites in developmental gene promoters are achieved. Comparison with other Rel–DNA complex structures leads to a general model for DNA recognition by Rel proteins.

Addresses: [1]European Molecular Biology Laboratory, Grenoble Outstation, BP 156, F-38042 Grenoble Cedex 9, France and [2]European Molecular Biology Laboratory, Meyerhofstraße 1, D-69117 Heidelberg, Germany.

Present Addresses: [†]Department of Structural Biology, Stanford University, Stanford, CA 94305–5400, USA, [‡]Chemistry Department, University of York, Heslington, York YO10 5DD, UK and [§]Department of Pathology, Colorado State University, Fort Collins, CO 80523–1671, USA.

*Corresponding author.
E-mail: mueller@embl-grenoble.fr

Key words: development, immune system, NF-κB, Rel protein, transcription factor

Received: 2 November 1998
Revisions requested: 18 December 1998
Revisions received: 18 March 1999
Accepted: 19 March 1999

Published: 29 June 1999

Structure July 1999, 7:841–852
http://biomednet.com/elecref/0969212600700841

 ISSN 0969-2126

Introduction

Rel proteins form a family of eukaryotic transcription factors that are involved in immune response and development in vertebrates and insects and in the activation of viruses such as human immunodeficiency virus 1 (HIV-1) and Herpes [1–7]. All family members share the Rel homology region (RHR), which is responsible for DNA binding, dimerization and nuclear localization. Rel proteins are divided into two classes: members of class II contain a transcription activation domain, whereas members of class I do not. Class I includes the mammalian proteins NF-κB p50 and p52, their precursors p105 and p100, respectively, and the *Drosophila melanogaster* protein Relish. Class II includes the mammalian proteins NF-κB p65, RelB and c-Rel, and the insect proteins Gambif1, from the human malaria vector *Anopheles gambiae* [8], and Dorsal and Dif, from *Drosophila*.

In insects, Rel proteins play a role in immunity and development. Their nuclear localization in *Drosophila* is controlled by the Toll–Cactus signalling pathway, which is related to the interleukin-1–IκB pathway in mammals [9–11]. An extracellular peptide ligand binds to and activates the Toll receptor [9,12] leading to phosphorylation of Cactus, a member of the IκB family [13], disruption of the Cactus–Rel protein complex and nuclear translocation of the Rel protein. Following this mechanism, Gambif1 is translocated to the nucleus in immune-responsive fat- body cells of *A. gambiae* upon bacterial challenge [8]. Gambif1 can bind to κB-like sequences and activate transcription of a reporter gene in co-transfection assays [8]. The RHR of Gambif1 shares 70% and 40% amino acid sequence identity with Dorsal and Dif, respectively, and 43%, 40% and 44% with human NF-κB p50, p52 and p65, respectively (Figure 1). Both Dorsal [12,14] and Dif [15] are involved in immune reactions in *Drosophila*, but show slightly different DNA-binding specificities.

Dorsal also acts as a developmental morphogen and its nuclear gradient establishes dorsoventral polarity in the early *Drosophila* embryo [16,17]. Nuclear localization of Dorsal on the ventral side of the embryo leads to activation of the genes *twist* (*twi*) and *snail* (*sna*) and to repression of the genes *zerknüllt* (*zen*) and *decapentaplegic* (*dpp*) [17]. There are two important determinants of the

threshold response to the Dorsal gradient: intrinsic binding-site affinity and interactions with other factors [18–21]. The promoters of Dorsal target genes contain multiple copies of binding sites that have varying affinity to the Dorsal protein. The affinity of the Dorsal sites to the protein sets the expression limit, and the overall number of sites determines the expression level [19].

Crystal structures of the homodimeric DNA complexes of NF-κB p50 [22,23], p52 [24] and p65 [25] have been reported. The crystal structure of the mouse p50/p65 heterodimer bound to DNA has also been determined [26]. The RHR folds into two β barrel immunoglobulin-like domains. Of these, the N-terminal specificity domain contains the 'recognition loop', which contacts DNA bases in the major groove, and the C-terminal domains of two Rel monomers form the dimer interface. Each dimerization domain contacts the DNA backbone via two loops.

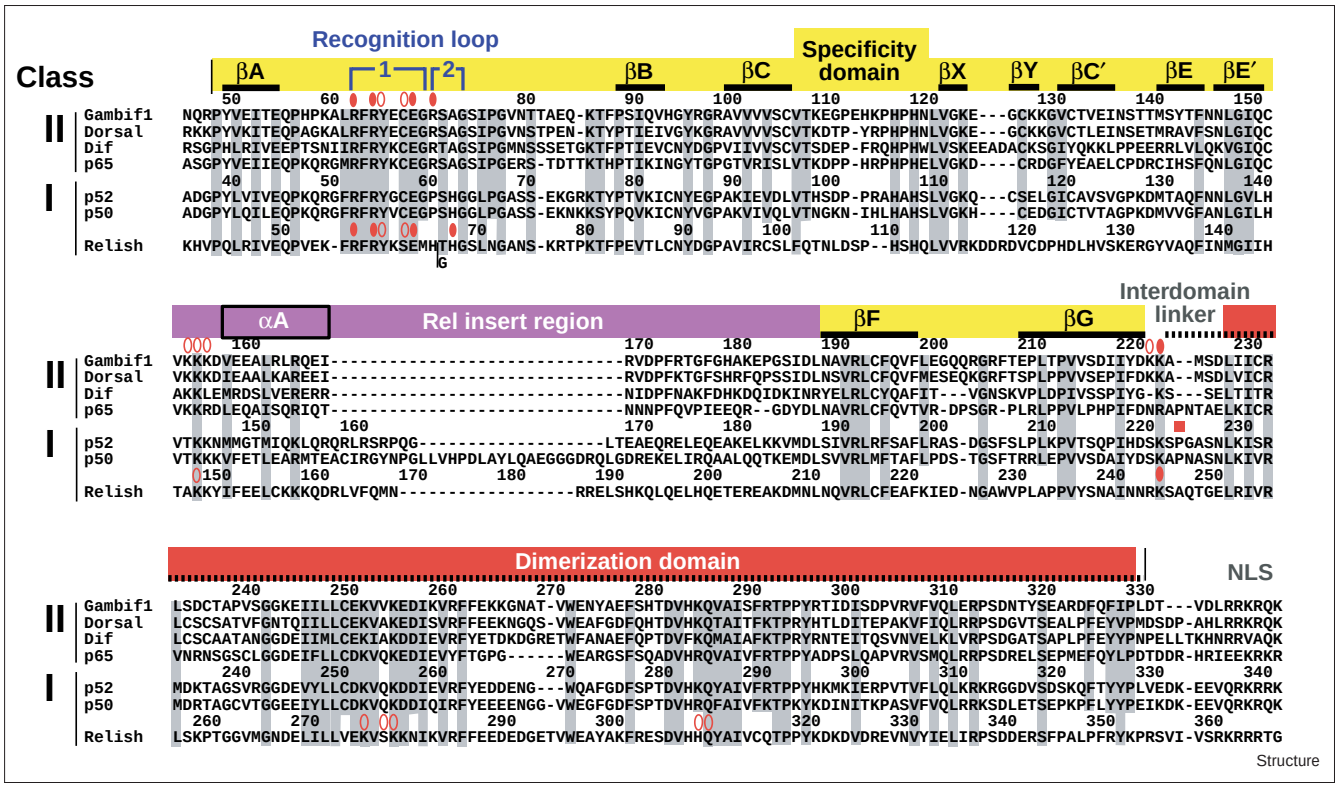
Here, we present the first X-ray structure of an insect Rel protein–DNA complex. The structure of the N-terminal specificity domains of the Gambif1 homodimer bound to DNA reveals a new base-specific contact to the outermost guanines of the undecameric DNA site, formed by an arginine residue that is conserved among all class II Rel proteins. The importance of this contact for binding affinity is demonstrated by electrophoretic mobility shift assays. Given that Gambif1 is the closest known homologue of Dorsal, the structure serves as a good model for Dorsal and explains differential Dorsal affinity to DNA sites in promoter regions of developmental target genes. Comparison with previously reported co-crystal structures suggests a general model for DNA recognition by Rel dimers.

Results and discussion

Structure solution and overall structure

The RHR of Gambif1 (Figure 1) was co-crystallized with a DNA duplex comprising a strong Dorsal-binding site derived from the *Drosophila* gene *zen* (Figure 2). The structure was solved using multiple isomorphous replacement (MIR) at 3.3 Å resolution using six different iodo-DNA derivatives (Table 1). The model was built based on the

Figure 1

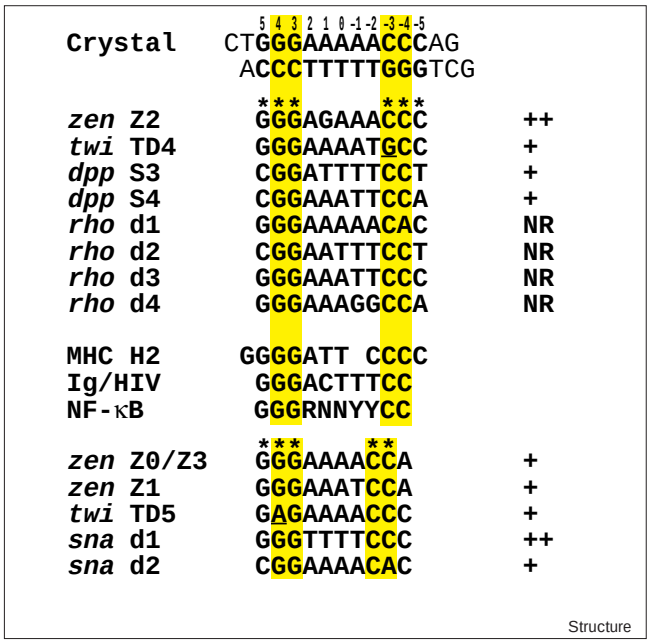


Sequence alignment of the RHR of human and insect Rel proteins. Amino acid numbering for Gambif1 and human NF-κB p52 is given above each sequence, numbering for human NF-κB p50 is below the sequence. The assignment of secondary structure elements (black bars, β strands; box, α helix) is based on the immunoglobulin convention as suggested for NFAT [28]. The Gambif1 construct is delineated by vertical lines; disordered regions are marked with a dashed bar. Residues of Gambif1 and human p50 contacting DNA bases and the DNA backbone are marked with red filled and open ovals, respectively. A proline in the linker that is conserved among mammalian Rel proteins is marked with a red square. Conserved residues are highlighted in grey.

NF-κB p52–DNA complex structure [24] and refined at 2.7 Å resolution (Table 2). It comprises the N-terminal specificity domain (residues 48–222), half of the 15-mer DNA duplex and 36 water molecules. The model has a crystallographic R factor of 21.9% (R_{free} = 28.8%) and shows excellent stereochemistry with 84.4% of the residues in the most favoured and none in the disallowed regions of the Ramachandran plot (calculated using the program PROCHECK [27]). Although the sequence of the DNA duplex used for co-crystallization deviates from strict twofold symmetry (Figure 2), a crystallographic twofold axis passes through the center of the DNA-binding site.

The N-terminal domain core of Gambif1 is a nine-stranded β barrel (Figures 3 and 4a). Its structure is very similar to that of human NF-κB p50 and p52 ($rms_{138C\alpha}$ is 1.2 Å, except residues 152–187) but it has a different position on the DNA: the domain is rotated by approximately 10° around the dimer dyad (anticlockwise in the view of Figure 3b). Residues 152–187 form the Rel insert region (RIR), a varying sequence element between β strands E′ and F (Figure 1). The RIR of Gambif1 folds into an α helix (αA, residues 157–167) and a long loop (residues 168–187). It resembles the RIR of NF-κB p65. Most of the RIR element packs against the domain core. The conserved residue Arg164 anchors helix αA as predicted [24]. In the RIR of Gambif1, helix αA is one turn shorter than in the RIR of NF-κB p50, and p52 and helix αB, which is present in NF-κB p50 and p52, is missing (Figure 4a). The reported structures suggest that helix αA is present in all Rel proteins, whereas helix αB is unique to class I. Residues corresponding to the RIR in the Rel-related protein NFAT interact with proteins bound to adjacent DNA sites [28]. Variability within the RIR might reflect different affinities for such partners.

Figure 2



Alignment of DNA sites. The top strand of the DNA duplex used in crystallization (crystal) and all other sequences are denoted in 5′→3′ direction. The binding site is shown in bold. The sites are sorted according to a potential five and four base-pair spacing between GG core elements (yellow). Several natural Dorsal sites from promoters of the *zen* [20], *twi* [38], *dpp* [37], *rho* [50] and *sna* [36] genes and sites for p50 and p52 homodimers (MHC H2), p50/p65 heterodimers (Ig/HIV) and the NF-κB consensus site (NF-κB) are shown. Guanines in the *zen* sites Z2 and Z0/Z3 that have been shown by methylation interference to be directly contacted by the protein [32] are marked with asterisks. Sites showing strong [32,36] and weaker [19,32,37] Dorsal affinity are marked with '++' and '+', respectively (NR, not reported). Exchange of the underlined guanine in *twi* TD4 to cytosine and the underlined adenine in *twi* TD5 to guanine increases Dorsal affinity [19].

The DNA is in a B-form-like conformation, is not bent and has an average twist of 9.7 base pairs/turn (as determined using the program CURVES [29]). In the crystal lattice, DNA duplexes are packed end-on along the z-axis and the overlapping bases form a Watson–Crick base pair. Because of the end-to-end stacking and the acentric position of the binding site in the DNA duplex (Figure 2), only one of the two possible orientations of the duplex can occur within a single continuous stack of DNA. In space group $P4_322$, in which a crystallographic dyad runs through the center of the binding site, stacks in the two possible orientations must be equally distributed throughout the crystal. The presence of this crystallographic dyad results in averaged electron density for the central non-palindromic base pairs ± 2, ± 1 and 0. The final $2F_o–F_c$ electron-density map is of high quality for the DNA backbone and the dyad-related bases and is consistent with a superposition of A:T base pairs in both orientations at the five central positions (Figure 4b). The 'averaging' around the dyad does not impair our conclusions concerning the protein–DNA interface because only base pairs ± 2 are contacted directly by the protein. The correct assignment of space group $P4_322$ was confirmed as described in the Materials and methods section.

Mobility of the dimerization module

Although the DNA and the N-terminal specificity domain are well defined in the experimental electron-density map, very little electron density is observed for the C-terminal dimerization domain such that it could not be modelled. There are a number of possible reasons for this disorder: first, proteolytic cleavage of the protein and the absence of the domain in the crystals; second, distinct domain positions in each monomer caused by the asymmetry of the DNA, resulting in weak density because of 'averaging' around the dyad; third, mobility resulting from dimer dissociation; and fourth, mobility of the dimerization module relative to the rest of the complex.

To exclude the first possibility, the presence of the C-terminal domains in the crystals was confirmed by

Table 1

Crystal structure solution.

Data set*	Native 1 (Phasing)	Native 2 (Refinement)	Iodo-1	Iodo-2	Iodo-3	Iodo-4	Iodo-5	Iodo-6	Native B
Resolution (Å)	20.0–3.3	20.0–2.7	20.0–3.1	20.0–3.5	20.0–3.1	20.0–4.0	20.0–4.2	20.0–3.5	20.0–3.1
ESRF beamline	BM14	ID02	BM14	BM14	BM14	BM14	BM14	BM14	ID02
Reflections observed/unique	53128/5937	33776/8927	35745/7041	19446/4543	25138/6610	11549/2823	9008/2909	15735/4567	41952/7228
R_{sym}† (%)	9.0 (31.8)	4.1 (28.6)	7.7 (27.3)	9.5 (28.0)	13.2 (26.9)	10.3 (22.2)	12.9 (29.9)	9.0 (24.9)	6.1 (33.1)
Completeness (%)	98.7 (90.5)	82.9 (88.5)	97.9 (98.3)	92.8 (94.9)	91.9 (89.0)	81.6 (75.2)	95.8 (92.9)	91.3 (93.5)	99.3 (98.6)
Phasing (Native 1)									
Isomorphous difference‡ (%)			15.4	18.2	22.0	12.7	15.6	18.5	
Phasing power§ acentric/centric		1.35/1.11	1.14/0.93	0.81/0.63	1.12/0.84	1.18/0.77	1.29/0.81		
R_{Cullis}# acentric/centric			0.77/0.67	0.83/0.73	0.90/0.89	0.79/0.80	0.79/0.83	0.79/0.80	
Overall figure of merit¶ acentric/centric/all		0.51/0.71/0.55							

*Native 1 and Native 2 contain the *zen*-derived DNA site ('crystal' in Figure 2). Derivatives iodo-1, iodo-2 and iodo-3 contained 5-iodocytosine at positions ± 5, ± 4 and ± 3, respectively, iodo-4 and iodo-5 contained 5-iodouracil at positions ± 1 and ± 2, respectively; iodo-6 contained 5-iodouracil at positions ± 1 and ± 2. Native B contains a DNA site of perfect symmetry except a central A:A mismatch (5′-GCTGGGAAATTCCCA-3′, binding site underlined) and shows 37.9% isomorphous difference to Native 2. Unit-cell axes are 88.4 Å × 88.4 Å × 95.8 Å. †$R_{sym} = \Sigma | Ii - \langle Ii \rangle | / \Sigma | \langle Ii \rangle |$, where Ii is the intensity of the individual reflection and ⟨Ii⟩ is the mean value of its equivalent reflections. Values given in parentheses correspond to the highest resolution shells. ‡Isomorphous difference = $\Sigma |F_{PH} - F_P| / \Sigma |F_{PH}|$, where F_{PH} and F_P are the derivative and native structure-factor amplitudes, respectively. §Phasing power is the mean value of the heavy atom structure-factor amplitudes divided by the lack of closure. #R_{Cullis} is the mean lack of closure divided by the mean isomorphous difference. ¶The figure of merit is defined as $\langle \Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha) \rangle$, where α is the phase and P (α) is the phase probability distribution.

protein gel analysis (data not shown). In accordance with this, the crystal packing permits placement of the domains. To test the second possibility, Gambif1 was co-crystallized with a symmetric DNA duplex containing only a central A:A mismatch (GCTGGGAAATTCCCA, the binding site is underlined). Gambif1 forms nearly isomorphous crystals with this DNA (Native B; Table 1). In this crystal form, the central base pairs show defined, non-averaged electron density, but the C-terminal domain could still not be located in difference Fourier maps calculated with model phases. Thus, the disorder does not result from DNA asymmetry. Furthermore, a lower symmetry space group as a possible reason for the poor density of the C-terminal domain was also excluded (see the Materials and methods section). The third possibility, dissociation of the DNA-bound protein dimer, is unlikely to be responsible given that Gambif1 forms stable homodimers in solution. Furthermore, other Rel proteins form stable dimers in their free and DNA-bound state and many interface residues are conserved throughout the Rel family [24]. We therefore conclude that the domain disorder is caused by the mobility of the entire dimerization module.

Table 2

Refinement statistics (Native 2).

Space group	$P4_322$
Unit-cell dimensions (Å)	87.6 × 87.6 × 96.2
Resolution (Å)	10.0–2.7
Number of protein residues	175
Number of nucleotides	15
Number of solvent molecules	36
Total number of non-hydrogen atoms	1717
Rms deviations from ideal geometry	
bond lengths (Å)	0.011
bond angles (°)	1.7
R_{cryst}/R_{free} (%)*	21.9/28.8

*R_{cryst} and R_{free} are calculated from reflections of the working set (8171 reflections) and the independent test set (404 reflections, 5%), respectively.

This mobility might arise from differences in the linker connecting the dimerization domains with the well-ordered specificity domains: the interdomain linker is two residues shorter in Gambif1 and Dorsal and three residues shorter in Dif than it is in mammalian Rel proteins (Figure 1). In particular, the insect proteins lack a proline residue in the linker sequence. In previously reported structures, this proline residue was shown to make important positioning contacts to the sugar–phosphate backbone. Mobility of the dimerization module with respect to the specificity domains might allow Gambif1 (and Dorsal) to bind to two types of DNA sites with different half-site spacing, leading to a broader specificity spectrum (see below).

Binding to tandem sites

Some promoters contain adjacent ('tandem') binding sites for Rel dimers. In the HIV-1 and *twi* promoters, centres of the Rel sites are 14 and 15 base pairs apart, respectively [19,30]. The packing observed in our crystals corresponds to the arrangement of two Rel dimers on the Dorsal tandem site in the *twi* promoter (Figure 5). Neighbouring dimers are located on opposite sides of the DNA helix, with the transcription activation domains facing opposite sides. The Rel dimers approach each other via their N-terminal domains through a loop (residues 80–86) immediately following the recognition loop (Figure 5a). There are, however, no specific dimer–dimer interactions. The sidechains of Glu83 and Gln84, which are both within contact distance to β strands X and E of the adjacent complex, are disordered. Indeed, cooperative binding to tandem sites has not been reported. For steric reasons, adjacent DNA duplexes are tilted away from the dimer–dimer contact point by an angle of 30° (Figure 5b). Accommodating both Rel dimers on a natural tandem site would, therefore, rely on bending of the DNA.

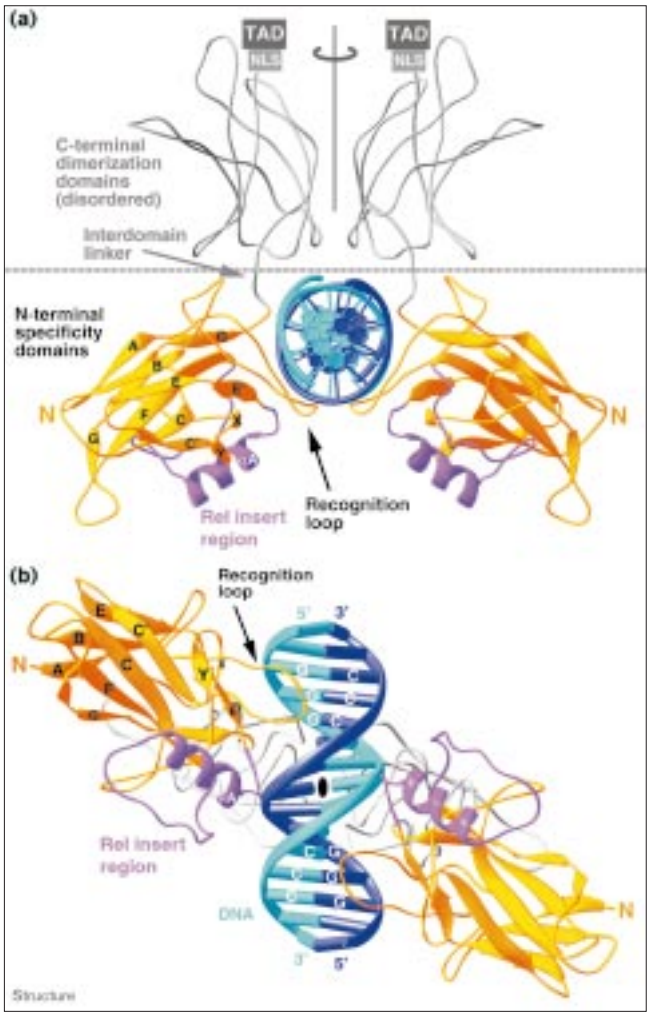
DNA recognition

A diagram of the observed Gambif1–DNA contacts is shown in Figure 6a. Gambif1 binds three contiguous guanines at positions ± 5, ± 4 and ± 3 in the major groove using a stack of three arginine sidechains presented by the recognition loop (Arg70, Arg64 and Arg62, respectively). The guanidinium groups of these arginine residues make bidentate interactions with the O6 and N7 hydrogen-bond acceptors of the guanines (Figure 6b). The sidechains of Arg62 and Arg64 are each buttressed by a structural water molecule that is hydrogen bonded to a backbone carbonyl. The sidechain of Glu68 contacts cytosine ± 3 and buttresses Arg64. Lys222 in the interdomain linker is the last residue included in the model and contacts adenine/thymine ± 2 with its sidechain. The contacted bases show average electron density. Both orientations of the A:T base pair ± 2 are found in natural sites (Figure 2) suggesting that Lys222 can contact thymine O4 or adenine N7. For one DNA orientation, the sidechains of Tyr65 and Glu68 pack against the thymine ± 2 methyl group. Tyr65 and Cys67 position the recognition loop through phosphate backbone contacts. A stretch of three lysine residues (153–155) at the N-terminal end of helix αA clamps the DNA at the central minor groove. Further, the sidechain of Lys221 contacts the guanine ± 4 phosphate.

Our observations are in good agreement with biochemical data for Dorsal. Selected and amplified binding (SAAB) analysis shows that the terminal G:C stretches of the Dorsal consensus site $GGG(T)_{4–5}CC$ are essential for high-affinity binding [31]. Methylation interference analysis demonstrated that the three guanines on either end of the *zen* site Z2 are contacted by Dorsal [32] (Figure 2). Although the center of the binding site (positions ± 2, ± 1, 0) can tolerate base substitutions, there is a clear preference for A:T base pairs [31]. This central part shows a narrow minor groove ($A0_{O4'}$–$T2_{O4'}$ = 5.8 Å) and large propeller twists. The sequence-dependent DNA conformation is sensed by the protein ('indirect readout') and occurs together with a widening of the major groove accommodating the recognition loops (Figure 3b).

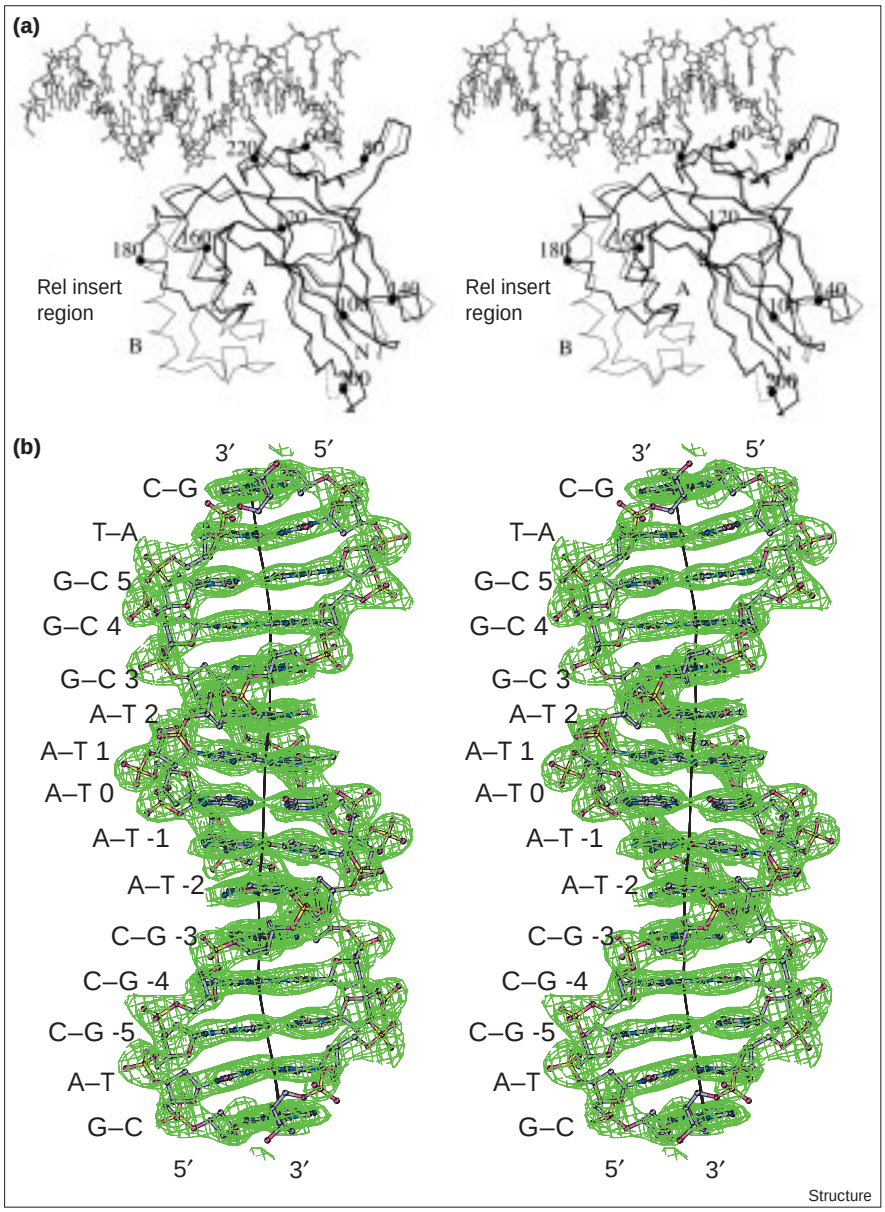
The specific contact of Arg70 to guanine ± 5 has not been observed in other Rel protein–DNA structures. The p65 homodimer and p50/p65 heterodimer structures contain a

Figure 3



Structure of the Gambif1 homodimer–DNA complex. (a) Overall view of the structure along the DNA helical axis with the dimer dyad vertical. The DNA is shown in blue. Colour coding for the N-terminal domain is as in Figure 1. The disordered C-terminal domains have been replaced by those of NF-κB p52 assuming the same relative orientation with respect to the N-terminal domains and are depicted in grey. The nuclear localization signal (NLS) at the C-terminal end of the RHR is followed by a transcription activation domain (TAD). (b) View of the complex structure from below as compared with (a). This figure was generated using the program RIBBONS [51].

Figure 4



(a) Stereoview Cα traces of the N-terminal domains of Gambif1 and human NF-κB p50 (thick and thin line, respectively) after least-squares superposition of Cα atoms of the domain core (omitting the RIR). Compared to Figure 3a, the model has been rotated by 90° around the dimer dyad. Loop FG is disordered in the p50 structure and is represented as a dashed line. The Cα atom of every 20th residue is drawn as a black sphere. This figure was generated using the program MOLSCRIPT [52]. (b) Stereoview of the final σ_A-weighted $2F_o–F_c$ electron density for the DNA contoured at 1.2σ. The view is from the top, onto the central major groove. The electron density for the five central non-palindromic base pairs ± 2, ± 1 and 0 is consistent with a superposition of A:T base pairs in both orientations. The depicted model represents a combination of two independently refined DNA strands (see Materials and methods section). The DNA helical axis was calculated with the program CURVES [29]. This figure was generated using the program BOBSCRIPT [53].

base-paired thymine and an adenine overhang, respectively, at position 5 of the p65 half-site [25,26]. The absence of guanine at this position explains why a corresponding Arg41–guanine-5 contact was not observed. To test the role of the novel contact in Gambif1 DNA affinity, we performed a competition binding analysis in an electrophoretic mobility-shift assay (EMSA). Exchange of guanine ± 5 in both half-sites with any other base strongly decreased the binding affinity (Figure 7a). To achieve 50% competition, a more than 100-fold excess of any double mutant competitor over the wild-type probe was needed. Thus, this previously unseen contact is important for affinity of Gambif1 to the DNA.

Differential DNA binding by Rel proteins of class I and II

The Rel recognition loop (62-RFRYECEGRSAG-73) is mainly responsible for specific DNA binding and can be divided into two regions (Figures 1 and 6c). Conserved region 1 (62-RFRYECEG-69) encompasses residues Arg62, Arg64 and Glu68 (underlined) which contact G:C base pairs ± 3 and ± 4. Corresponding residues in p50, p52 and p65 make similar contacts (Figure 8). Region 2 differs between class II (70-RSAG-73) and class I (PSHG) Rel proteins. Residue Arg70 (underlined) is conserved only within class II and contacts guanine ± 5 in the Gambif1–DNA structure. Members of class I bind to this guanine with a histidine (His67/62 in human p50/p52;

Figures 6c and 8). To allow these disparate contacts, region 2 adopts a different conformation in Gambif1 and human p50/p52. All known Rel proteins can be classified according to these sequence motifs, except Dif and Relish. In Dif, only a Ser→Thr substitution is present in region 2 (RTAG). In Relish, residue changes and an insertion are present (RFRYKSEM-HGTHG); region 2 comprises two histidines (underlined) that could contact DNA.

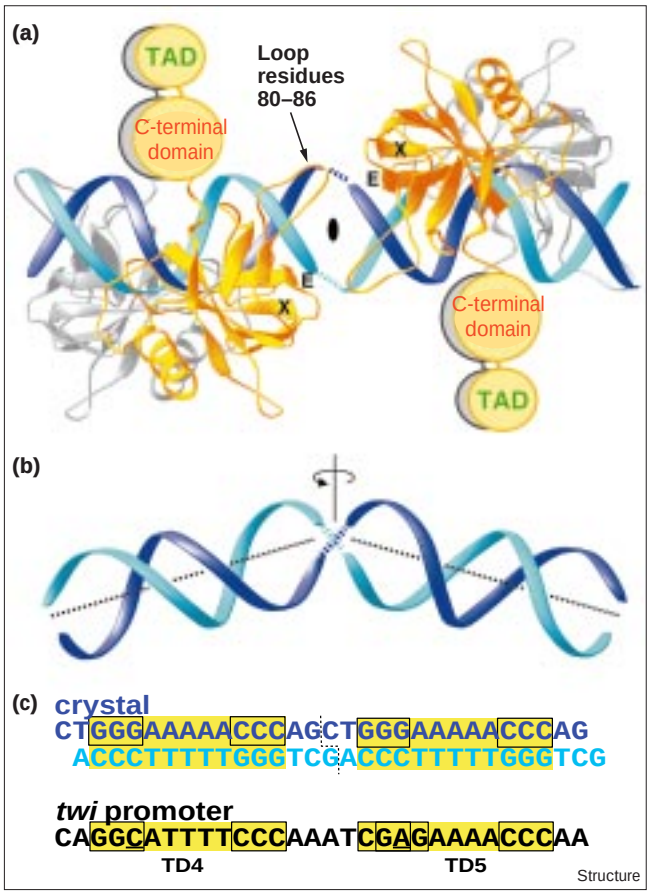
The importance of region 2 for subclass specificity was shown by site-directed mutagenesis [33]. Replacement of Arg41 and Ala43 in region 2 of p65 by the corresponding residues of p50 (Pro and His) led to a mutant protein binding efficiently to a p50-selective site [33]. EMSA competition analysis revealed the importance of the Arg70–guanine ± 5 contact in region 2 of Gambif1 for DNA affinity. In similar experiments, the corresponding contact of class I proteins has been tested for p50 (His67–guanine ± 5) [34]. Changing one of the outermost guanines of the specific DNA site MHC H-2 (Figure 2) to another base caused a 7–20-fold drop in affinity [34]. Thus, Rel homodimers of both classes bind preferably to guanines at positions ± 5 of undecameric sites, albeit with different residues. In contrast, the consensus sites obtained by selection techniques are decameric [31,35]. In such selection experiments, however, cytosine (and hence guanine in the contacted strand) is most frequently found at the eleventh position (position –5) [31,35].

A general model for DNA recognition by Rel proteins

Each DNA half-site has a core GG element (base pairs ± 4, ± 3) that is contacted by a Rel monomer with two conserved arginines in region 1 of the recognition loop (Figure 8). Further, a conserved glutamate of region 1 binds to cytosine ± 3 in the opposing strand and forms buttressing hydrogen bonds with one or more residue(s) (Figures 6c and 8). Considering the wide sequence divergence within the Rel family, these contacts are remarkably conserved; they are even partly found in NFAT [28]. Sites missing a guanine in one of the core elements can still allow binding but with lower affinity than if guanine is present (Figure 2, and see below). Changes in affinity or specificity can also result from changes in contacts at positions flanking the core element. The inner base pair ± 2 is contacted by a non-conserved residue in the interdomain linker (Figures 1 and 8), whereas the outer base pair ± 5 can be contacted by histidine (class I) or arginine (class II) in region 2 of the recognition loop (Figures 6c and 8).

A five base-pair spacing between GG core elements is observed in all known structures except the mouse p50–DNA complex in which the elements are six base pairs apart [23]. A DNA site with a six base-pair spacing used for co-crystallization with the p65 homodimer, however, allowed for specific interactions in only one half-site [25], demonstrating that the spacing is too wide.

Figure 5

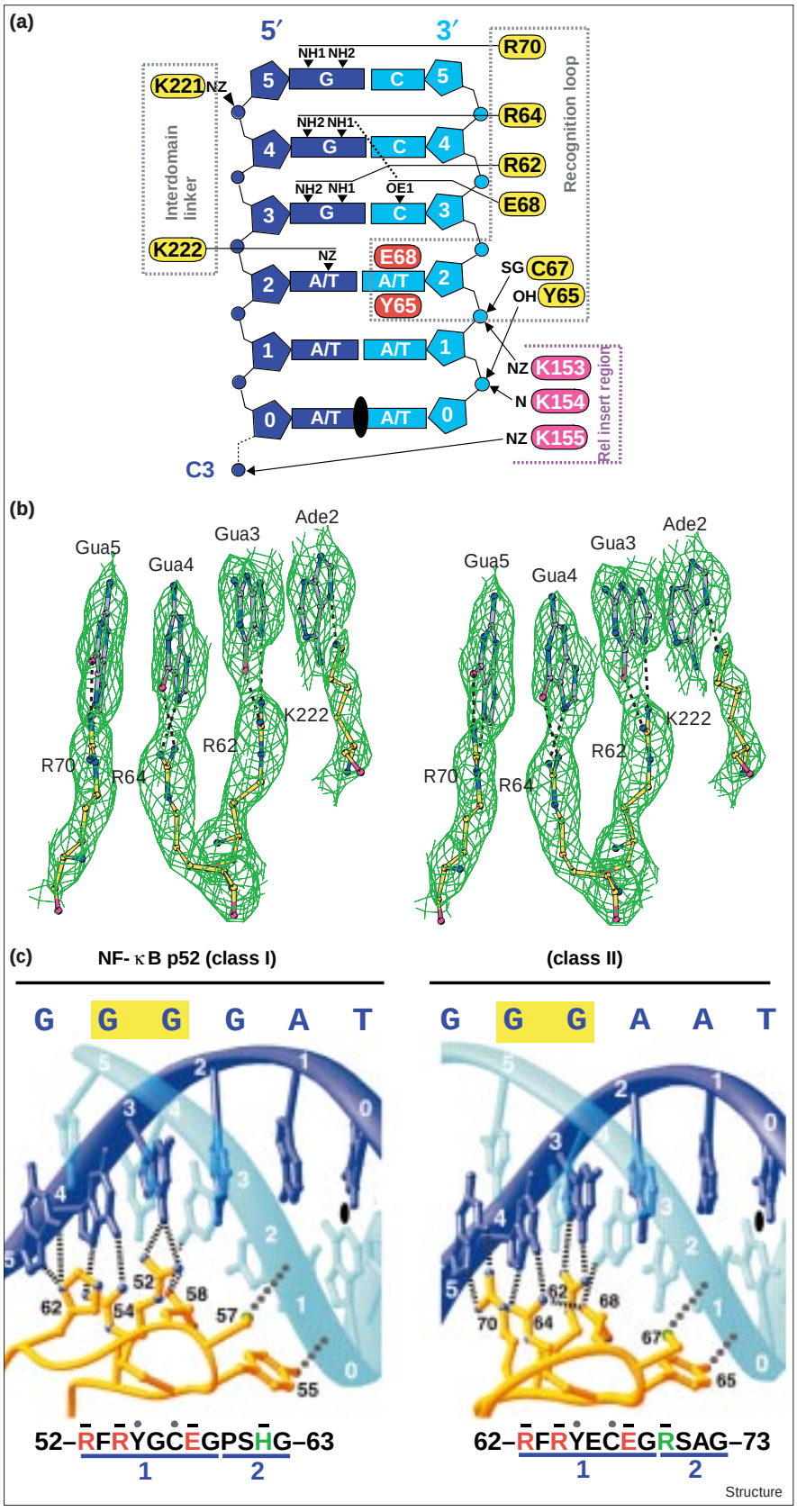


A model for binding of Rel dimers to tandem sites. (a) Crystal packing of two adjacent Gambif1–DNA complexes. TAD, transcription activation domain. (b) View of two adjacent DNA duplexes from the top as compared with (a). Parts (a) and (b) were generated using the program RIBBONS [51]. (c) Alignment of the DNA sequence in the crystal with the natural tandem site sequence from the *twi* promoter.

There is evidence that Gambif1 and Dorsal also bind to decameric sites with a four base-pair spacing (Figure 2). First, methylation interference showed that Dorsal contacts corresponding guanines in the *zen* site Z0/Z3 [32] (Figure 2). Second, EMSA competition analysis showed that Gambif1 binds with even higher affinity to an idealized decameric site with a four base-pair spacing than to a undecameric site with a five base-pair spacing (Figure 7b), suggesting that all arginine–guanine contacts are maintained. In insect Rel proteins, the shorter interdomain linker lacking the positioning proline and the mobility of the dimerization module relative to the specificity domains might allow a spacing of four or five base pairs.

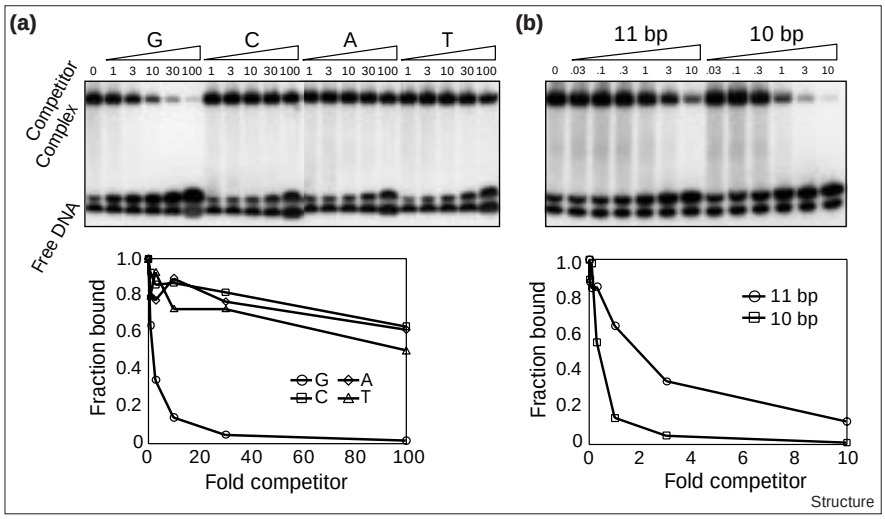
What causes preferred orientation of Rel heterodimers on asymmetric sites? In the p50/p65 heterodimer complex structure, p50 and p65 bind to the 5′ and the 3′ half-site, respectively, of the Ig/HIV site 5′-GGGACTTTCC-3′ [26].

Figure 6



DNA recognition by Gambif1. (a) Diagram of polar (yellow and magenta) and van der Waals (red) interactions between one protein monomer and a DNA half-site. (b) Stereoview of base-specific interactions in the DNA major groove. The final σ_A-weighted $2F_o–F_c$ electron density is contoured at 1.3σ. The DNA backbone is omitted for clarity. This figure was generated using the program BOBSCRIPT [53]. (c) Comparison of DNA recognition by NF-κB p52 (class I) and Gambif1 (class II). Polar base and backbone contacts of residues in the recognition loop with one half-site are shown as dashed and dotted lines, respectively. Guanines at position 5 are bound by a histidine or an arginine sidechain in region 2 of the recognition loop of p52 and Gambif1, respectively. In the sequences below, contact residues of region 1 and region 2 are shown in red and green, respectively. Base and backbone contacting residues are marked with bars and dots, respectively. This figure was generated using the program RIBBONS [51].

Figure 7



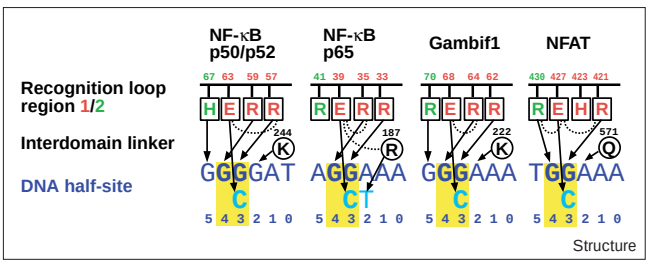
Competition analysis of Gambif1–DNA binding. Electrophoretic mobility shift assays (top) and quantification of bands (bottom) were carried out as described in the Materials and methods section. The amount of competitor DNA is given as multiples of labelled DNA above each lane (onefold, 33 pmol). Oligodeoxynucleotides were self-annealed to give symmetric 15-mer DNA duplexes with a central A:A mismatch. The free DNA consists of an active and an inactive fraction (upper and lower bands, respectively). The inactive fraction did not vary substantially with competitor concentration and might correspond to hairpins formed by single DNA strands. A DNA fragment of unrelated sequence did not show any competition in the concentration range examined (data not shown). **(a)** Testing of the specific contact Arg70–guanine ±5. The sequence of the wild-type probe (G) was 5′-CTGGGAAATTCCCAG-3′ (binding site underlined). Sequences of the double mutant competitors were 5′-CTCGGAAATTCCGAG-3′ (C), 5′-CTAGGAAATTCCTAG-3′ (A), and 5′-CTTGGAAATTCCAAG-3′ (T; mutant positions underlined). **(b)** Comparison of symmetric undecameric and decameric sites with a half-site spacing of five and four base pairs, respectively. The undecameric site probe G and a decameric site probe (5′-TCTGGGAATTCCCAG-3′; binding site underlined) were used.

In contrast, our results suggest that p65 could also bind a third guanine using Arg41; this residue corresponds to Arg70 of Gambif1 that contacts guanine ± 5. Other factors such as specific contacts of the linker residues (Figure 8) or indirect readout of the asymmetric central region of the DNA must influence the orientation of the heterodimer.

Structural basis for differential Dorsal site affinity

Knowing the structural determinants of site affinity helps in understanding the threshold response to the Dorsal gradient in the *Drosophila* embryo. A literature survey reveals that strong Dorsal binding to promoter sites requires intact GGG half-site elements [32,36] (Figure 2). Base alterations within these elements lead to reduced affinity [19,32,37]. The Gambif1–DNA structure shows that the loss of specific arginine–guanine contacts is responsible for this. Differences in promoter site affinity to Dorsal can determine whether or not a gene will respond at a given concentration of Dorsal. For example, the *zen* gene is repressed in lateral regions of the *Drosophila* embryo by low Dorsal levels that are insufficient to activate the *twi* gene [16]. Consistent with this, Dorsal sites in the *twi* promoter show significantly lower affinity for Dorsal than the sites in the *zen* promoter [21,38] (Figure 2). The *twi* sites contain base changes in either one of the GGG elements. The Gambif1–DNA structure predicts that these alterations would disrupt guanine interactions with the arginine residues of region 1 of the recognition loop, leading to a lower affinity. Indeed, exchange of the differing bases in *twi* sites to guanine leads to increased affinity [19].

Figure 8



Comparison of DNA recognition by various Rel proteins and NFAT. Specific contacts made by residues in the recognition loop (boxes) and the interdomain linker (circle) are indicated by arrows for human p50 [22] and p52 [24], mouse p65 [26], Gambif1 (this work) and NFAT [28]. The conserved glutamate in region 1 of the recognition loop forms different buttressing hydrogen bonds (dashed lines). The GG core half-site elements are highlighted in yellow.

Biological implications

The NF-κB/Rel family of transcription factors regulates the expression of specific genes involved in immunity and development in vertebrates and insects. In response to their activation by a variety of external stimuli, Rel protein dimers translocate to the nucleus and bind in a sequence-specific manner to DNA sites in target gene promoters. Understanding differential gene expression in response to NF-κB activation requires the structural and energetic characterization of the protein–DNA interactions involved.

The *Drosophila* Rel protein Dorsal is a developmental morphogen and its nuclear gradient establishes dorsoventral polarity in the early *Drosophila* embryo.

The affinity of Dorsal for DNA sites in target promoters is an important determinant of the response to this gradient. Gambif1 from *Anopheles gambiae* is a close homologue of Dorsal with 70% amino acid sequence identity in the Rel homology region. Its translocation to the nucleus in fat-body cells in response to bacterial infection has been demonstrated.

Our crystal structure of the N-terminal specificity domain of Gambif1 bound to DNA is the first structure of an insect Rel protein–DNA complex to be determined. The N-terminal β-barrel domain closely resembles the domains of the mammalian Rel proteins NF-κB p50, p52 and p65. However, the positioning of this domain with respect to the DNA and the structure of the insert region, a sequence element varying within the Rel family, are different. The C-terminal dimerization domains are poorly ordered. Most probably this results from the absence of important protein–DNA backbone contacts involving the linker between the N-terminal and the C-terminal domains. This linker is characteristically shorter in insect Rel proteins than in mammalian homologues. Increased flexibility between the N-terminal and the C-terminal domains might be of functional importance in allowing the Rel protein to adapt to the different half-site spacings found in natural Dorsal sites. The packing of adjacent complexes in the crystal corresponds to the arrangement of tandem sites in the HIV-1 and *twi* promoters.

Our crystal structure reveals a base-specific contact at the outermost base pair of the undecameric DNA site, in which an arginine residue contacts a guanine base. The frequent occurrence of a guanine base in natural target sites at this position and electrophoretic mobility-shift assays suggest the functional importance of this contact. It had not been observed in earlier crystal structures of Rel protein–DNA complexes and explains how differences in DNA site affinity of Dorsal are achieved. A stack of three arginine residues within the so-called 'recognition loop' binds a GGG half-site element. Natural DNA sites missing a guanine in these elements show lower affinity for Dorsal than sites with guanine, resulting from the loss of a specific arginine–guanine contact. These differences in site affinity can cause differential biological responses, which are important at different stages of development.

On the basis of previous structures and our results, we suggest a general model for DNA recognition by Rel protein dimers. A core GG element in each half-site (XGGY) is contacted by three conserved residues, two arginines and a glutamate, in the recognition loop of a Rel monomer. This core element is flanked by two base pairs. The inner position (Y) is contacted by a non-conserved residue in the interdomain linker. At the outer position (X), guanine is present in many cases and is contacted differently by members of the two Rel subclasses: members of class I use a histidine whereas members of class II can use an arginine. This difference might account for subclass-specific DNA recognition.

Materials and methods

Sample preparation

Gambif1 was co-crystallized with DNA following a strategy worked out for human NF-κB p52 [39]. A clone encoding residues 48–341 and an N-terminal histidine tail was truncated to a gene encoding residues 48–329 by site-directed mutagenesis as described [39]. The obtained construct corresponds to the minimal RHR of Gambif1. The protein was overexpressed in *Escherichia coli* BL21 (DE3) cells at 23°C for 14 h after induction with 0.1 mM IPTG. Cells were resuspended in purification buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 5 mM DTT, 2 mM EDTA, 0.5 mM PMSF, 1 μg/ml pepstatin A and 1 μg/ml leupeptin) and lysed by sonication. After subsequent polyethylene imine and ammonium sulfate precipitation (final concentration 55% saturation), Gambif1 was purified to homogeneity by cation-exchange and size-exclusion chromatography (SP-Sepharose and Superose-12, Pharmacia). Pure protein formed a stable dimer as judged by size-exclusion chromatography. A protein sample was characterized by N-terminal automated Edman degradation and electrospray ionization mass spectroscopy. Oligodeoxynucleotides were synthesized chemically, purified by anion exchange chromatography (Mono-Q HR 10/10, Pharmacia) and dialyzed against distilled water. After lyophilizing, they were redissolved in demineralized water and the concentration was adjusted to 20 mg/ml. Protein–DNA complexes were reconstituted by mixing solutions of 475 μM Gambif1 homodimer in buffer A (20 mM HEPES pH 7.5, 0.15 M NaCl, 3 mM DTT) and 1.1 mM annealed DNA duplex in a molar ratio of 1.2:1.

Co-crystallization, data collection and processing

Crystals were grown by vapor diffusion at 21°C using the hanging-drop method. The reservoir solution contained 2–5% PEG 400, 50 mM MES pH 5.6, 5 mM $MgCl_2$ and 3 mM DTT. For hanging drops, 2 μl complex solution was mixed with 2 μl reservoir solution. Small crystals appeared over night but large crystals took at least two weeks to grow to a maximum size of $200 \times 200 \times 400\ \mu m^3$. Crystals containing iodo-oligonucleotides were significantly smaller than those that did not contain iodo-oligonucleotides. Crystals were harvested in reservoir solution containing 8% PEG 400. The concentration of PEG 400 was increased stepwise to 32% and the crystals were frozen at 100K in a stream of nitrogen using an Oxford cryosystem. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble from single crystals using image plate (MAR Research) and CCD detectors (ESRF Grenoble). Data were processed with programs DENZO and SCALEPACK [40]. Using intense ondulator radiation at ESRF beamline ID02 crystals diffract to 2.5 Å resolution in the direction of the DNA helical axis but more weakly in the other directions. Strong diffuse scattering was observed, most probably resulting from the disordered domain. Crystals belong to space group $P4_322$ with unit-cell dimensions a=b= 87.6 Å, c= 96.2 Å and a solvent content of 51% (V_M = 2.50 $Å^3$/Da [41]). Native crystals were not completely isomorphous to each other and therefore no attempt was made to combine data sets of different native crystals. Plots of the cumulative intensity distribution of the data did not indicate crystal twinning.

To confirm the assigned space group $P4_322$ with a dyad running through the center of the DNA-binding site, native data and iodo-DNA derivative data were reprocessed in space group $P4_3$ lacking the dyad, which resulted in similar R_{merge} values despite lower redundancy. Two independent tests were performed. First, the model was expanded to space group $P4_3$ and rigid-body refined in $P4_3$ against Native 1, the data set used for phasing and most isomorphous with the iodine derivatives. Phases were calculated and used in a difference Fourier synthesis of Native 1 with iodouracil derivatives iodo-4, iodo-5 and iodo-6, which yielded peaks of equal height at iodine positions related by a

noncrystallographic dyad in space group $P4_3$ (Table 3). This demonstrates that the two possible orientations of the DNA duplex are equally occupied and that the situation in the crystal is best described by space group $P4_322$. (Alternatively the crystal can also be described in space group $P4_3$ but the DNA would still have to be modelled in two possible orientations. Derivatives iodo-1, iodo-2 and iodo-3 were not suitable for this analysis because in these derivatives the two chemically different oligonucleotide strands contain substitutions at dimerically related positions.) Second, a model expanded to space group $P4_3$ and missing the central non-palindromic A:T base pairs was refined against data processed in space group $P4_3$. For this analysis we used a native data set extending only to 2.9 Å resolution but with higher redundancy. Standard and simulated-annealing omit electron-density maps showed average density for the omitted bases confirming the space group assignment. Furthermore, no additional electron-density features were detected in the region of the C-terminal domains, which rules out the possibility that the C-terminal domains have been 'averaged out' as a result of a wrongly assigned space group.

Structure solution and refinement

Initial molecular replacement trials using various search models based on the human NF-κB p50 and p52 complex structures were unsuccessful. Therefore, iodo-DNA derivatives were used for structure solution with the MIR method. Iodine positions in derivatives iodo-1, iodo-2 and iodo-3 were located in difference Patterson maps using data set Native 1 (Table 1). Difference Fourier maps then revealed iodine positions in derivatives iodo-4, iodo-5 and iodo-6. Program MLPHARE [42] was used for refinement of heavy-atom parameters and calculation of MIR phases. The resulting MIR electron-density map at 3.3 Å resolution allowed us to manually place the NF-κB p52 N-terminal domain core and a 15-mer duplex of canonical B-DNA. Very weak, uninterpretable electron density was present in the expected region for the C-terminal domain. Solvent flattening was not successful probably because of the large space occupied by the disordered domain. The MIR map allowed tracing of the RIR (residues 152–187) and loop FG using program O [43]. Several loops were rebuilt and sidechains differing between NF-κB p52 and Gambif1 were replaced and adjusted in regions with clear experimental electron density and they were omitted in poorly ordered regions.

At this stage data set Native 2 became available and was subsequently used during the refinement in the resolution range 10.0–2.7 Å. Rigid-body refinement of the model dropped the R factor from 0.50–0.46 (resolution range 10–6.0 Å) and was followed by maximum-likelihood refinement implemented in the program REFMAC [44]. The quality of the σ_A-weighted $2F_o–F_c$ and difference Fourier maps was further improved by iterative updating of the partial model with the automated refinement procedure ARP [45,46]. Although map improvement with ARP is mostly used at a resolution of 2.5 Å or higher, the procedure was successful with our data, probably because of the high quality of the starting model and the favourable data-to-parameter ratio resulting from the domain disorder. Data were anisotropically scaled with the program REFMAC, although no uniform bulk-solvent correction was applied. Truncated sidechains were stepwise reintroduced during refinement and 36 well-defined water molecules were added. At final stages, the program X-PLOR [47] and improved stereochemical DNA parameters [48] were used for refinement. Temperature factors were individually refined using bond restraints on temperature factors of 2.5 $Å^2$ for protein mainchain and 3.5 $Å^2$ for protein sidechain and DNA atoms. The average temperature factor of the final model is 52.7 $Å^2$. Two DNA strands with half occupancy (sequence I: 5′-CTGGGAATTTCCCAG-3′, sequence II: 5′-CTGGGTTAAACCCAG-3′) approximate the situation in the crystal where A:T and T:A base pairs superimpose at the five non-palindromic positions ± 2, ± 1 and 0 (see above). The good refinement statistics (Table 2) probably result from the use of the 2.1 Å resolution structure of NF-κB p52 as a reference during the refinement and the improved data-to-parameter ratio because of the disordered domain. They further indicate that the Bragg scattering contribution from this domain is negligible.

Table 3

Peak heights in difference Fourier synthesis of the iodouracil derivatives.

Derivatives	Iodo-4	Iodo-5	Iodo-6
Peak in base pair 2		10	11
Peak in base pair 2*		10	9
Peak in base pair 1	11		9
Peak in base pair 1*	10		9
Peak in base pair –1	12		8
Peak in base pair –1*	10		8
Peak in base pair –2		11	10
Peak in base pair –2*		11	12

In the crystal, two orientations of the DNA duplex are observed. Base pairs in orientation I are denoted following the numbering in Figure 2 and the same base pair numbering with an asterisk is used for base pairs in orientation II. Peak heights are given in number of standard deviations above the mean of the difference Fourier map calculated to the resolution limits of the derivatives as given in Table 1. For iodo-4 with iodouracil substitutions in base pairs ± 1 and iodo-5 with substitutions in base pairs ± 2 we observe four difference peaks per derivative, iodo-6 contains iodouracil substitutions in four positions (base pairs ± 1 and ± 2) giving rise to eight difference peaks. Peak heights in base pairs without asterisks are to be compared with peaks related by a noncrystallographically dyad in base pairs with asterisks. They become crystallographically equivalent in space group $P4_322$. (For example iodo-4: 11σ peak in base pair 1 (orientation I) compared to 10σ peak in base pair 1* (orientation II).) The differences between peak heights related by the noncrystallographic dyad are small and randomly distributed. The largest difference is observed between related peaks in base pairs 2/2* of iodo-6 (11σ/9σ). The analysis demonstrates that the two possible orientations of the DNA duplex are equally occupied in the crystal and confirms the assigned space group $P4_322$.

Competition binding analysis

A 15-mer oligodeoxynucleotide G (5′-CTGGGAAATTCCCAG-3′) containing an undecameric binding site (underlined) was radioactively labelled with 170 μCi [γ-^{32}P]ATP according to Silberklang *et al.* [49]. After purification on a 15% denaturing (8 M urea) polyacrylamide gel, the labelled probe was annealed with an excess of unlabelled oligodeoxynucleotide in 50% buffer A to give a concentration of 100 μg/ml. Binding reactions were carried out in a volume of 20 μl. First, 3 μl 100 μg/ml radioactive probe (33 pmol) and 3 μl competitor DNA in 1–100-fold excess (for sequences, see the legend of Figure 7) were premixed in 10 μl binding buffer (20 mM HEPES pH 7.5, 5% glycerol, 1 mM EDTA, 100 mM NaCl, 1 mM DTT, 0.05% Triton X-100, 1 mg/ml BSA). The reaction was started by the addition of 4 μl 3.13 μM (12.5 pmol) freshly prepared Gambif1. Only about half the total amount of DNA was active in protein binding. After incubation at 25°C for 20 min, reactions were subjected to electrophoresis in 0.25 × TBE on 6% native polyacrylamide gels (45 min at 6 W). The gels were transferred to Whatman paper, dried at 75°C for 40 min and autoradiographed. The amount of radioactivity was quantified by scintillation counting of excized bands. To obtain the fraction of labelled probe bound by Gambif1, the counts per minute (cpm) of the complex bands were divided by the cpm of complex formed in the absence of competitor.

Accession codes

The coordinates have been deposited with the Brookhaven Protein Data Bank under accession code 1bvo and will be released upon publication of this article.

Acknowledgements

We thank EMBL/ESRF Joint Structural Biology Group members A Thompson and G Leonard for help at the ESRF beamline BM14 and

J Grimes and B Rasmussen for help at the ESRF beamline ID02. We also thank M Wiersma and R Eritja (EMBL Heidelberg) for oligodeoxynucleotide synthesis, A Perrakis for advice on refinement and C Petosa for comments on the manuscript.

References

1. Baeuerle, P.A. & Henkel, T. (1994). Function and activation of NF-κB in the immune system. *Annu. Rev. Immunol.* **12**, 141-179.
2. Baeuerle, P.A. & Baltimore, D. (1996). NF-κB: ten years after. *Cell* **87**, 13-20.
3. Baldwin, A. (1996). The NF-κB and IκB proteins: new discoveries and insights. *Annu. Rev. Immunol.* **14**, 649-681.
4. Chytil, M. & Verdine, G.L. (1996). The Rel family of eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **6**, 91-100.
5. Thanos, D. & Maniatis, T. (1995). NF-κB: a lesson in family values. *Cell* **80**, 529-532.
6. Kanegae, Y., Tavares, A.T., Belmonte, J.C.I. & Verma, I.M. (1998). Role of Rel/NF-κB transcription factors during the outgrowth of the vertebrate limb. *Nature* **392**, 611-614.
7. Bushdid, P.B., *et al.,* & Kerr, LD (1998). Inhibition of NF-κB activity results in disruption of the apical ectodermal ridge and aberrant limb morphogenesis. *Nature* **392**, 615-618.
8. Barillas-Mury, C., *et al.,* & Kafatos, FC. (1996). Immune factor Gambif1, a new rel family member from the human malaria vector, *Anopheles gambiae*. *EMBO J.* **15**, 4691-4701.
9. Belvin, M.P. & Anderson, K.V. (1996). A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12**, 393-416.
10. Medzhitov, R. & Janeway, C., Jr. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**, 295-298.
11. Gonzalez-Crespo, S. & Levine, M. (1994). Related target enhancers for dorsal and NF-κB signaling pathways. *Science* **264**, 255-258.
12. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M. & Hoffmann, J.A. (1996). The dorsoventral regulatory gene cassette *spätzle*/*Toll*/*cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973-983.
13. Gilmore, T.D. & Morin, P.J. (1993). The IκB proteins: members of a multifunctional family. *Trends Genet.* **9**, 427-433.
14. Lemaitre, B., *et al.*, & Hoffman, JA. (1995). Functional analysis and regulation of nuclear import of dorsal during the immune response in *Drosophila*. *EMBO J.* **14**, 536-545.
15. Ip, Y.T., *et al.,* Levine, M. (1993). Dif, a dorsal-related gene that mediates an immune response in Drosophila. *Cell* **75**, 753-763.
16. Roth, S., Stein, D. & Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**, 1189-1202.
17. St Johnston, D. & Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-219.
18. Jiang, J., Cai, H., Zhou, Q. & Levine, M. (1993). Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. *EMBO J.* **12**, 3201-3209.
19. Jiang, J. & Levine, M. (1993). Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741-752.
20. Kirov, N., Zhelnin, L., Shah, J. & Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in *Drosophila*. *EMBO J.* **12**, 3193-3199.
21. Thisse, C., Perrin-Schmitt, F., Stoetzel, C. & Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila twist* gene by the dorsal gene product. *Cell* **65**, 1191-1201.
22. 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.
23. Ghosh, G., van Duyne, G., Ghosh, S. & Sigler, P.B. (1995). Structure of NF-κB p50 homodimer bound to a κB site. *Nature* **373**, 303-310.
24. Cramer, P., Larson, C.J., Verdine, G.L. & Müller, C.W. (1997). Structure of the human NF-κB p52 homodimer–DNA complex at 2.1 A resolution. *EMBO J.* **16**, 7078-7090.
25. Chen, Y.Q., Ghosh, S. & Ghosh, G. (1998). A novel DNA recognition mode by the NF-κB p65 homodimer. *Nat. Struct. Biol.* **5**, 67-73.
26. Chen, F.E., Huang, D.B., Chen, Y.Q. & Ghosh, G. (1998). Crystal structure of p50/p65 heterodimer of transcription factor NF-κB bound to DNA. *Nature* **391**, 410-413.
27. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
28. Chen, L., Glover, J.N., Hogan, P.G., Rao, A. & Harrison, S.C. (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**, 42-48.
29. Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63-91.
30. Jones, K.A. & Peterlin, B.M. (1994). Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**, 717-743.
31. Pan, D. & Courey, A.J. (1992). The same dorsal binding site mediates both activation and repression in a context-dependent manner. *EMBO J.* **11**, 1837-1842.
32. Ip, Y.T., Kraut, R., Levine, M. & Rushlow, C.A. (1991). The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in Drosophila. *Cell* **64**, 439-446.
33. Coleman, T.A., Kunsch, C., Maher, M., Ruben, S.M. & Rosen, C.A. (1993). Acquisition of NFκB1-selective DNA binding by substitution of four amino acid residues from NFκB1 into RelA. *Mol. Cell Biol.* **13**, 3850-3859.
34. Urban, M.B. & Baeuerle, P.A. (1991). The role of the p50 and p65 subunits of NF-κB in the recognition of cognate sequences. *New Biol.* **3**, 279-288.
35. Kunsch, C., Ruben, S.M. & Rosen, C.A. (1992). Selection of optimal κB/Rel DNA-binding motifs: interaction of both subunits of NF-κB with DNA is required for transcriptional activation. *Mol. Cell Biol.* **12**, 4412-4421.
36. Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K. & Levine, M. (1992). Dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1518-1530.
37. Huang, J.D., Schwyter, D.H., Shirokawa, J.M. & Courey, A.J. (1993). The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. *Genes Dev.* **7**, 694-704.
38. Jiang, J., Kosman, D., Ip, Y.T. & Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev.* **5**, 1881-1891.
39. Cramer, P. & Müller, C.W. (1997). Engineering of diffraction-quality crystals of the NF-κB p52 homodimer–DNA complex. *FEBS Lett.* **405**, 373-377.
40. Otwinowski, Z. & Minor, W. (1997). Processing of diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
41. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
42. CCP4 (1994). Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-776.
43. Jones, A.T., Zhou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
44. Murshudov, G.M., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240-255.
45. Lamzin, V.S. & Wilson, K.S. (1997). Automated refinement for protein crystallography. *Methods Enzymol.* **277**, 269-305.
46. Perrakis, A., Sixma, T.K., Wilson, K.S. & Lamzin, V.S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D* **53**, 448-455.
47. Brünger, A.T. (1992) X-PLOR Version 3.1. A System for X-ray Crystallography and NMR (Yale University Press, New Haven, CT, USA.
48. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A.T. & Berman, H.M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D* **52**, 57-64.
49. Silberklang, M., Gillum, A.M. & RajBhandary, U.L. (1977). The use of nuclease P1 in sequence analysis of end group labeled RNA. *Nucleic Acids Res.* **4**, 4091-4108.
50. Ip, Y.T., Park, R.E., Kosman, D., Bier, E. & Levine, M. (1992). The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1728-1739.
51. Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958-961.
52. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
53. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.