DNA binding by an amino acid residue in the C-terminal half of the Rel homology region

Jia Liu[1], Qing R Fan[1], Mikiko Sodeoka[1], William S Lane[2], and Gregory L Verdine[1]*

[1]Department of Chemistry and [2]Microchemistry Facility, Harvard University, Cambridge, MA 02138, USA

Background: Members of the Rel family of transcription factors are important in regulating the inflammatory, acute phase, and immune responses in mammals. The structural basis for sequence-specific binding by Rel proteins is poorly understood, however. In the studies reported here, a new photoaffinity labeling procedure has been used to probe DNA contacts established by a Rel protein, the p50 homodimer of NF-κB.

Results: Using a novel post-synthetic modification method, 8-azido-2'-deoxyadenosine (N_3dA), a photoactive analog of 2'-deoxyadenosine, was introduced at a specific site within a consensus DNA binding site for the NF-κB p50 homodimer. Upon irradiation with ultraviolet light, the N_3dA-substituted DNA was efficiently photocrosslinked to p50. The crosslinked amino acid of p50 was identified as K244, which lies in the carboxy-terminal half of the Rel homology region (RHR). Mutation of K244 exerts strong effects on DNA binding, confirming the importance of this residue for p50–DNA interactions.

Conclusions: We have used N_3dA-containing DNA to identify a residue of NF-κB p50 that contacts DNA illustrating the value of this approach in studies of protein–DNA interactions. K244 appears not to contact a DNA base, but rather a phosphate moiety that lies nearby. The segment of protein sequence in which K244 resides has been implicated in dimerization. The results presented here suggest that the DNA-binding domain extends farther toward the carboxy-terminus than previously thought.

Chemistry & Biology September 1994, **1**:47–55

Key words: 8-azido-2′-deoxyadenosine, NF-κB, protein–DNA photocrosslinking, Rel transcription factors

Introduction

NF-κB is a transcription factor that is important in regulating the inflammatory, acute phase, and immune responses [1,2]. Its intracellular targets are DNA sequence elements that control the transcription of a number of different genes. In addition to regulating the expression of endogenous genes, NF-κB has been implicated as an obligate host factor in the activation of pathogenic viruses including adenovirus, cytomegalovirus, and human immunodeficiency virus (HIV-1). NF-κB is a heterodimer of two DNA-binding subunits, p50 and p65, both of which are members of the Rel transcription factor family.

Rel proteins contain a domain of ~280 amino acids, the Rel homology region (RHR), which mediates both DNA binding and dimerization. Dimers can be formed from two identical subunits or from two different Rel family members. In certain members of the Rel family, including p50 and several closely related proteins, the RHR is hyphenated by an insert of ~30 amino acids. Previous evidence has suggested that this hyphenation may mark the boundary between the functional subdomains of the RHR; all DNA contact residues identified to date have been localized in the amino-terminal half (RHR-A) [3–8], and the residues responsible for dimerization reside in the carboxy-terminal half (RHR-B) [8–10]. Despite these advances, the ability to reconcile structure with function in Rel proteins has been hampered by: (i) their lack of sequence similarity to any of the well-known families of structurally characterized DNA-binding domains [11,12]; (ii) their inability to tolerate truncation to any fragment smaller than the entire RHR [7,13]; and (iii) the lack of a high-resolution structure of any Rel family member.

Residues involved in sequence-specific DNA binding by Rel proteins have been identified by photocrosslinking using the NF-κB p50 homodimer [5]. Two residues of p50, Y60 and H67, were found to be in the contact interface between DNA and p50 [5]. Although site-specific photocrosslinking has proven useful for identifying DNA-contact residues of NF-κB and several other proteins [5,14,15], the generality of the approach is limited by several technical factors, of which the most problematic are inefficient photochemical crosslinking and lack of availability of reagents suitable for characterizing contacts to all four bases in DNA. Considering only those examples in which the photoprobe is placed in the intimate recesses of the protein–DNA interface, presently available nucleoside derivatives only allow the analysis of major groove contacts to T (using 5-bromo-2'-deoxyuridine [5,15,16], 5-iodo-2'-deoxyuridine [17], or 4-thio-2'-deoxyuridine [18]); and G (using 6-thio-2'-deoxyguanosine [18]). All of these reagents use free radicals generated by irradiation to produce crosslinks. These are typically difficult to form and react poorly with many of the functional groups of proteins. The recent development of conditions for efficient laser photolysis of 5-iodinated

*Corresponding author.

 ISSN 1074-5521

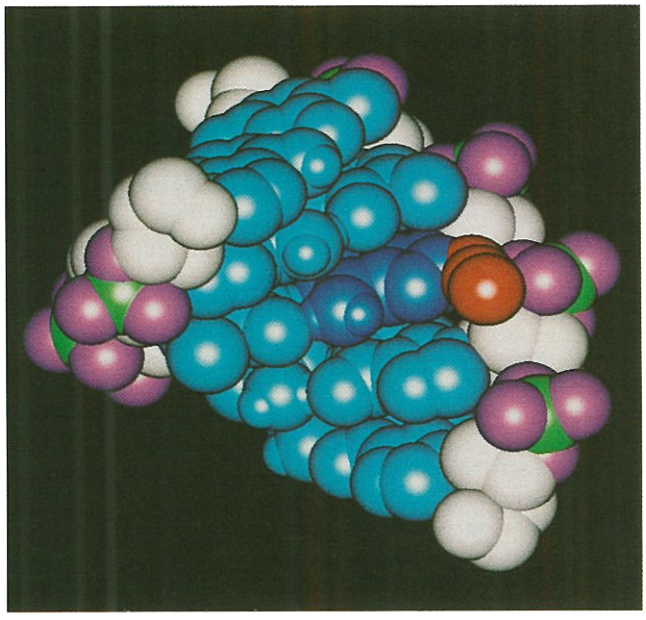

Fig. 1. Model of a DNA segment containing N_3dA. A five base-pair segment of model duplex DNA, 5'-d(GGATTC), in which A = N_3dA, is shown in van der Waals surfaces. The azide function is in red; the A base in dark blue; the remainder of the bases in light blue; O and P of the backbone phosphates in purple and green, respectively; remainder in white.

pyrimidines is a significant breakthrough, leading in some cases to highly efficient crosslinking [17].

Nitrenoid species (a nitrogen atom with six valence-shell electrons), which are generated through photolysis of azido-nucleosides, are an attractive alternative to radical photochemistry. The six-electron intermediates in nitrenoid photocrosslinking reactions are considerably more electron-deficient (electrophilic) than their radical counterparts and hence tend to be much more reactive and less selective [19]. The azidonucleosides N_3dA [20] and 5-azido-2'-deoxyuridine [21] have been incorporated enzymatically into DNA and shown to crosslink proteins with high efficiency. It has not been possible to position these photoprobes specifically at known sites, however, and it has therefore been impossible to use azide-based photocrosslinking to identify individual contact amino acid residues at specific sites.

The studies described here were aimed at developing a chemical method to introduce an azidonucleoside, N_3dA, into a specific site on DNA via solid-phase synthesis methods. The azide function of N_3dA resides in the major groove close to the DNA backbone (Fig. 1); while this position itself appears not to be an important contact site for proteins, it lies close by the *N7* position, a common protein contact site [11,22]. Thus, while N_3dA is not expected to disrupt specific DNA binding by proteins, it should be able to crosslink efficiently to nearby amino acid residues. Indeed, nonspecifically incorporated N_3dA has been shown to undergo efficient photocrosslinking to the Tet repressor [20]. Here we describe the site-specific incorporation of N_3dA into a consensus binding site for the p50 homodimer and its use in identifying a DNA contact residue of p50, Lys244 (K244). Surprisingly, K244 is located more than 175 residues carboxy-terminal to any DNA contact residue previously identified, and is well into RHR-B. These findings indicate that either the dimerization subdomain of the RHR directly contacts DNA, or the DNA-binding domain extends through all of RHR-A and the insert, including part of RHR-B.

Results

Attempted synthesis of N_3dA phosphoramidite

The most direct route for site-specific introduction of N_3dA into DNA would be to construct an N_3dA monomer appropriately equipped for solid-phase DNA synthesis, then to use this reagent in conventional automated oligonucleotide assembly. To test the feasibility of this route, we synthesized N_3dA, protected it by N^6-benzoylation and 5'-tritylation, then attempted to react the protected azidonucleoside with the phosphitylating reagent 2-cyanoethyl tetraisopropyl phosphordiamidite. Rather than react at the free 3'-OH, the phosphitylating reagent attacked the azide group exclusively, resulting in a phosphazide byproduct instead of the desired phosphoramidite [23]. These results indicate that the azide functionality is incompatible with solid-phase DNA synthesis using phosphoramidite chemistry. Furthermore, since the alternative H-phosphonate chemistry also involves P(III) compounds, this route is also expected to be incompatible with azido-nucleosides. We therefore turned our attention towards the development of a method for introduction of the azide function after solid-phase oligonucleotide synthesis.

Post-synthetic incorporation of the azide functional group into DNA

We have reported a method for the site-specific incorporation of 8-bromo-2'-deoxyadenosine (BrdA) into oligonucleotides using solid-phase synthesis chemistry [24]. The conversion from 8-bromoadenine to 8-azidoadenine is well-known in simple nucleosides [25,26]; hence, we attempted to carry out a similar transformation in DNA. Simple oligonucleotides (for example, d(T-BrdA-T)) were treated with azide ion under a variety of conditions, and the conversion of BrdA to N_3dA was monitored by diode array high-pressure liquid chromatography (HPLC); the presence of N_3dA in the desired product was detected by its characteristic UV spectrum (λ_{max}=282 nm, ε_{282nm}=1.07x10^4). Although in nucleoside model studies anhydrous solutions of NaN_3 (saturated) or $nBu_4N^+N_3^-$ (75 mM) in dimethylformamide (DMF) gave nearly quantitative conversion of BrdA to N_3dA upon prolonged heating (J.L. and G.L.V., unpublished results), these conditions yielded unacceptably high amounts of byproducts with simple oligonucleotides as substrates. Upon screening a variety of azide transfer conditions, we found that heating the simple oligonucleotides in anhydrous triethylammonium azide ($Et_3NH^+N_3^-$) in DMF at 75 °C for 8 h yielded oligonucleotides having

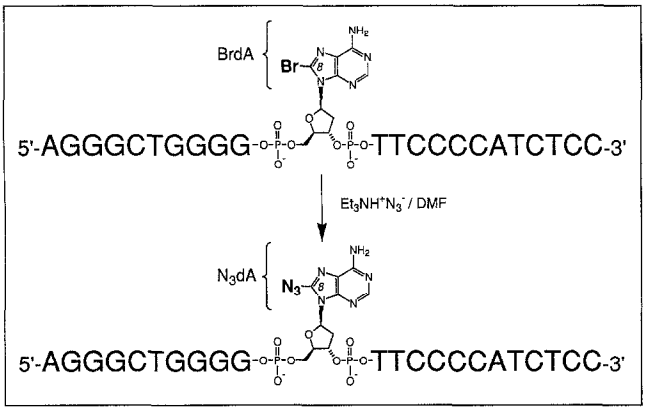

Fig. 2. Conversion of BrdA to N_3dA in a 23-mer oligonucletide through an azide transfer reaction.

up to 60 % N_3dA at the desired position. The principal byproduct in these reactions, 8-oxo-dA, was presumably formed as the result of direct attack on BrdA by residual H_2O or by an indirect process involving attack of Et_3N followed by hydrolysis during workup; under neutral conditions both BrdA and N_3dA are hydrolytically stable. Taking into account the 10 % 8-oxo-dA present as a contaminant in the starting BrdA-containing oligonucleotide, the azide transfer reaction proceeds with an efficiency approaching 70 %.

The optimized azide transfer conditions were then used to synthesize a 23-mer oligonucleotide (Fig. 2) containing a single N_3dA residue within the NFκB site of the MHC class I enhancer [27,28]; this sequence (5'-GGGGATTCCCC) is a consensus binding site for the NF-κB p50 homodimer [29]. In a typical displacement reaction, 80–100 % of the starting 23-mer was recovered following workup, as judged by UV quantification of the product. The nucleoside composition of the oligonucleotide product was confirmed by digestion with snake venom phosphodiesterase and alkaline phosphatase, followed by HPLC analysis (Fig. 3). In addition to the four naturally occurring deoxynucleosides, the digests revealed the presence of N_3dA and the byproduct 8-oxo-dA. In all solvent systems tested, N_3dA and BrdA elute too close together to allow accurate determination of their relative proportions directly from the ratio of peak areas. However, the UV spectra of the two are sufficiently distinct that the nucleoside composition could be determined using a computerized 'peak purity' routine. By this analysis, and taking into account differences between the extinction coefficients of BrdA, N_3dA and 8-oxo-dA, we determined that more than 95 % of the starting BrdA is consumed in the displacement reaction, yielding up to 70 % of N_3dA and as little as 30 % of 8-oxo-dA. No attempts were made to optimize the reaction conditions further; however, it may be that more rigorous exclusion of H_2O during the displacement reaction will give higher yields of the desired azidonucleotide.

In any event, the presence of 8-oxo-dA can be considered negligible for most cases, because proteins that make a contact to dA are expected strongly to prefer N_3dA over 8-oxo-dA; although N_3dA has the same arrangement of hydrogen bonding elements as dA, the arrangement in 8-oxo-dA is altered (the *N*7-H of 8-oxo-dA is a donor, whereas the *N*7 of dA is an acceptor).

Photo-crosslinking reactions

To address the efficiency of p50 crosslinking to azide-containing DNA, the 23-mer product of azide displacement was labeled with ^{32}P and annealed to an unlabeled complementary strand to generate a duplex molecule containing a single N_3dA residue (**A**):

5'-^{32}P-d(AGGGCTGGGG**A**TTCCCCATCTCC)-3'
3'-d(TCCCGACCCCTAAGGGGTAGAGG)-5'

The duplex 23-mer probe was then incubated with a recombinant fragment of human p50 (rp50), which contains residues 1–366 and binds DNA in much the same way as the wild-type protein [30]. The protein–DNA complex was exposed to UV light, and the products analyzed by denaturing PAGE. Irradiation of the rp50–oligonucleotide complex resulted in the formation of a new band having retarded gel mobility (Fig. 4, lane 2); the oligonucleotide alone yielded no such band (Fig. 4, lane 1), suggesting that the new band was the result of DNA crosslinking to the protein rather than to itself. Moreover, the crosslinked band was susceptible to competition by a 40-fold excess of unmodified specific oligonucleotide (Fig. 4, lane 4) but was unaffected by non-specific competitor (Fig. 4, lane 3). Under similar conditions, the single-strand N_3dA-containing 23-mer failed to yield a crosslinked product, as did duplex molecules containing BrdA and others having N_3dA at other positions in the 23-mer (data not shown). These data show that the crosslinking reaction is specific with respect to cognate binding of rp50, position in the sequence, duplex versus single-stranded DNA, and structure of the photoprobe.

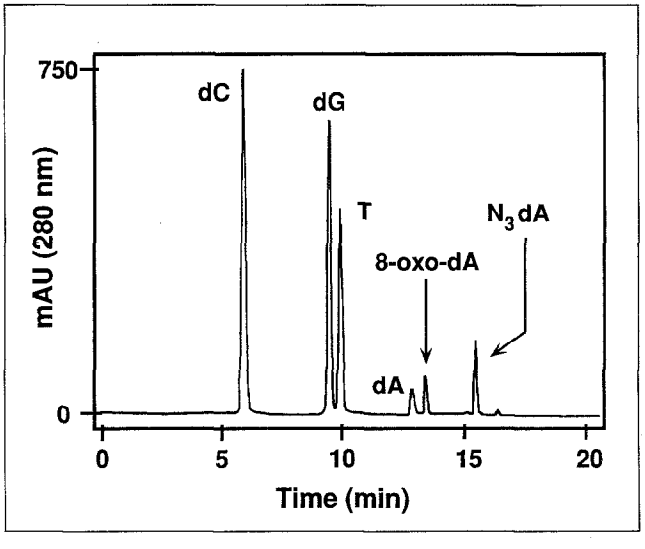

Fig. 3. Nucleoside composition analysis of the products resulting from azide displacement on a BrdA-substituted 23-mer.

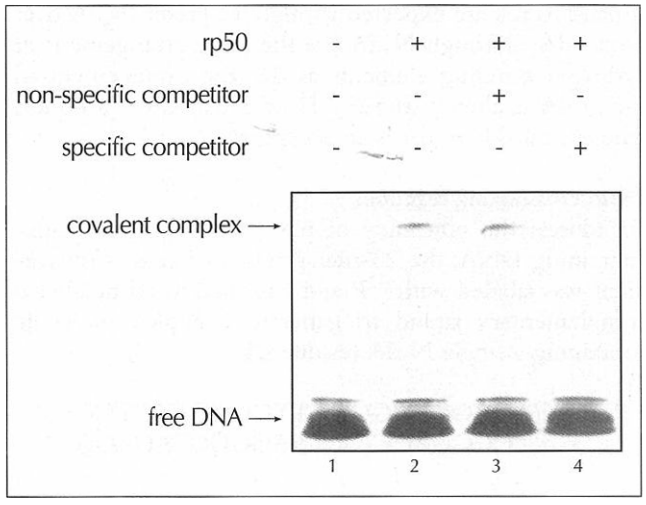

Fig. 4. SDS-PAGE analysis of photo-crosslinking reactions between rp50 and DNA. Lane 1, control crosslinking reaction without rp50; lane 2, reaction without competitor DNA; lane 3, reaction with a non-specific competitor DNA added; lane 4, reaction with a specific competitor DNA added.

Identification of crosslinked amino acids

To identify the amino acid residue of p50 that is close to the N_3dA moiety (Fig. 5), we carried out a large-scale photocrosslinking reaction. After protein–DNA incubation and UV irradiation, the reaction mixture was washed extensively against a urea-containing buffer to denature complexes that were not covalently bound and to desalt the mixture before chromatography. The mixture was fractionated by anion-exchange fast protein liquid chromatography (FPLC) chromatography to separate the covalent protein–DNA complex from unreacted DNA and protein. The complex was then reduced, alkylated, and exhaustively digested with trypsin. Fractionation of the digestion mixtures by anion-exchange FPLC yielded a late-eluting peak with a UV spectrum characteristic of DNA (λ_{max} ~260 nm). The material in this peak was precipitated and amino acid analysis showed that significant amounts of amino acids were present (data not shown). Peptide sequencing by automated Edman degradation gave a 21 amino acid sequence, LEPVVSDAIYDSXAPNASNLK. The crosslinked peptide precisely matches residues 232–252 of human p50 [13,31], and the single undetermined residue X in the sequenced peptide was thus unambiguously identified as K244 (Fig. 5). A second large-scale crosslinking reaction was digested with lysyl endopeptidase and sequenced, yielding DAIYDSXAPNASNLK, which corresponds to residues 237–252 of p50; again, K244 was found to be modified. Based on these data, we conclude that the photoactivated N_3dA residue was selectively crosslinked to K244 of rp50.

Mutational analysis

To explore the role of K244 in the p50–DNA interaction further, three variants of rp50 were made, K244R, K244A and K244M, each having a different amino acid substitution (Arg, Ala and Met, respectively) at position 244. The affinity of each of these proteins for a specific 23-mer oligonucleotide was quantified by the electrophoretic mobility shift assay (EMSA) (Fig. 6). To examine whether K244 might participate in electrostatic interactions with the DNA backbone, we performed EMSA analysis over a range of salt concentrations. Wild-type rp50 exhibits maximal DNA-binding efficiency at 100 mM NaCl, binding less strongly as the salt concentration is either increased or decreased. The K244R mutant protein, which retains positive charge in its side chain, exhibits similar preferences, but is more sensitive to high salt and

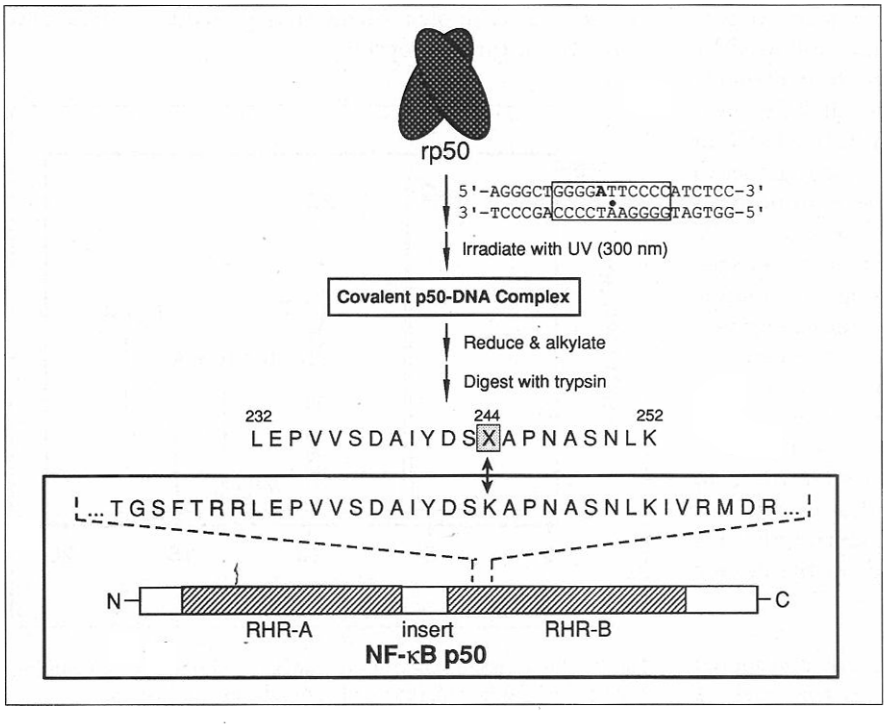

Fig. 5. Procedure used for identification of a DNA contact residue in rp50. A denotes N_3dA residue, • represents the center of pseudodyad symmetry.

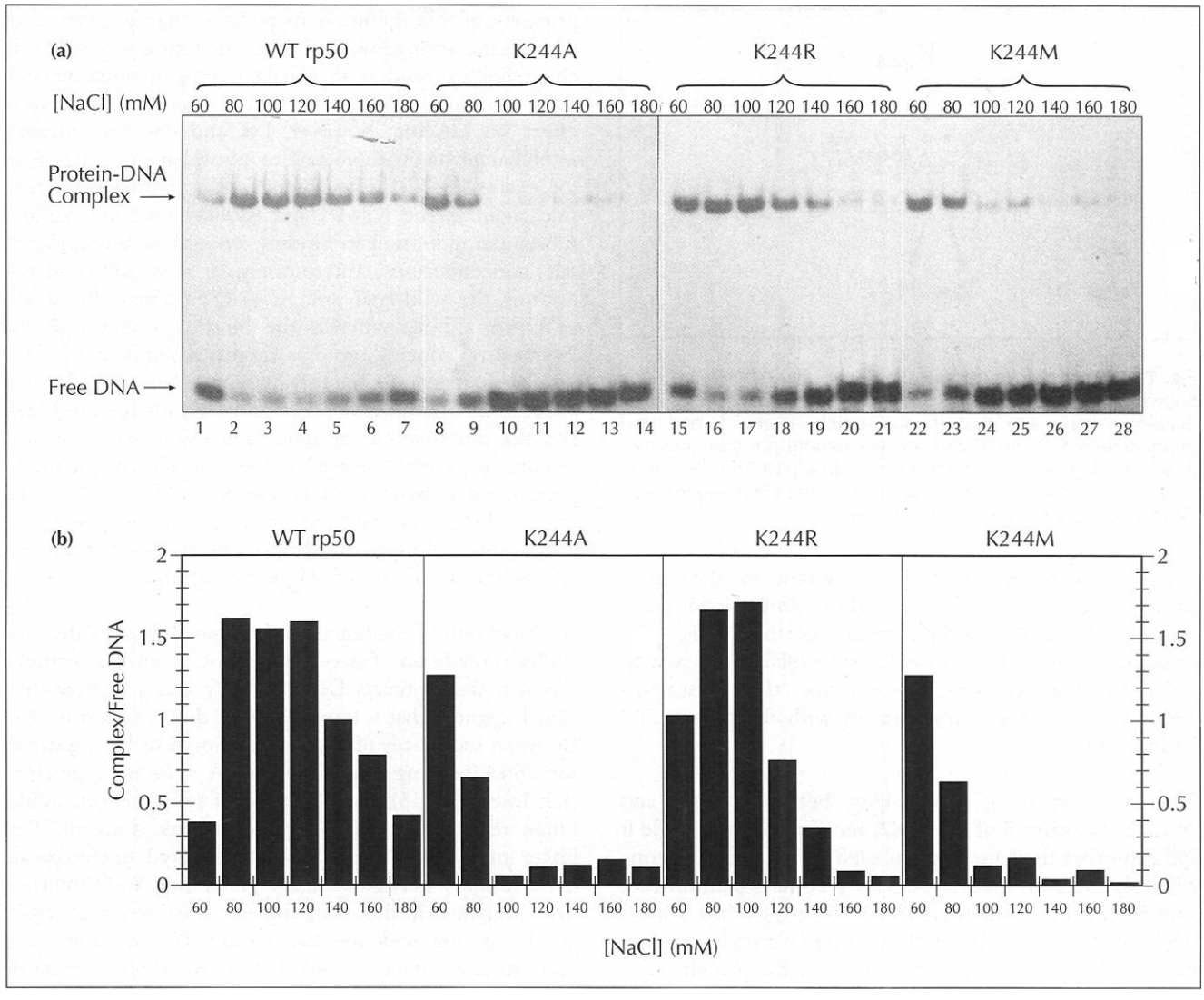

Fig. 6. DNA binding of wild-type rp50 and position 244 mutant proteins. **(a)** EMSA showing the effect of increasing salt concentration on DNA binding by wild type rp50 and the rp50 mutants K244A, K244R and K244M. **(b)** Histograms of the data given in (a) quantified by phosphor image analysis.

less sensitive to low salt than wild type. The two mutant proteins that have neutral side chains, K244A and K244M, behave very differently. These mutants bind much less tightly than either wild type or K244R at physiological salt concentrations (100–150 mM NaCl). Interestingly, the neutral side-chain mutants bind DNA more tightly than either wild-type or K244R at salt concentrations below the physiological level (60 mM NaCl). All these mutant proteins formed stable dimers, as judged by their mobility during gel permeation chromatography (data not shown). These data clearly show that the residue at position 244 has a strong effect on DNA binding, especially at physiological salt concentrations.

Discussion

In this work, we have developed a procedure for identifying amino acid residues that closely approach adenines in the major groove of DNA. The procedure relies on the site-specific introduction of a photoactive dA analog, N_3dA [20], into the protein–DNA interface. The nitrenoid intermediate generated by photolysis of N_3dA inserts itself more or less indiscriminately into nearby bonds; if these bonds are part of a protein, the resulting product is a covalent protein–DNA crosslink. The position of crosslinking is identified through proteolytic digestion, Edman microsequencing, and comparison with the known protein sequence. Here we have applied this method to identify DNA contact residues used by p50, a member of the Rel family of transcription factors.

The results described here reveal that K244 lies in close proximity to the adenine at position 5 of the Rel recognition element. In earlier work, we used the photoprobe 5-bromo-dU to identify Y60 and H67 as being proximal to base-pairs 4 and 5, respectively [5] (Fig. 7). Taken together, these results provide the remarkable demonstration that photoprobes positioned just a few Å apart in DNA crosslink amino acid residues that are separated by more than 175 amino acids in primary sequence. Thus, we

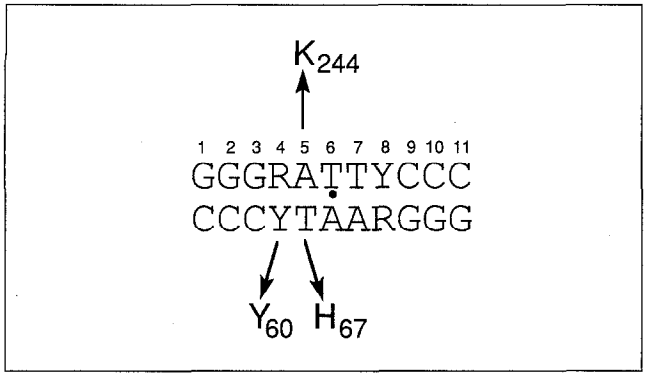

Fig. 7. Summary of photo-crosslinking results reported to date. Shown is a typical recognition element for Rel proteins, a pseudo-dyad-symmetric site that is purine-rich at the 5'-end and pyrimidine-rich at the 3'-end. Arrows denote positions in DNA at which photoactive nucleosides crosslink p50 and the amino acid residues that become crosslinked. In the DNA sequence, Y designates a pyrimidine and R a purine.

conclude that residues that are distant in the linear polypeptide chain are folded in three-dimensional space to form the DNA-binding surface of the RHR. This unusual feature of Rel proteins may explain in part why, unlike most known transcription factors, they resist truncation to a DNA-binding domain with fewer than 100 amino acids.

The observation of crosslinking between K244 and N_3dA at position 5 of the DNA recognition site, while it indicates that they are close to each other, does not conclusively establish that the two are a contact pair. Indeed, substitution of N^6-methyl-dA for dA at position 5 has a negligible effect on the binding of p50, strongly suggesting that A5 is not contacted directly by the protein (C.J. Larson and G.L.V., unpublished results). On the other hand, the binding data presented here on proteins mutated at position 244 show that K244 contributes substantially to the strength of the protein–DNA interaction, most likely by making a direct contact to DNA. These seemingly disparate results can be reconciled by proposing that K244 contacts not A5, but the phosphate group on the 5'-side of A5. Cross-linking between the dA nitrene and a Lys residue interacting with its 5'-phosphate is made plausible by the fact that the 8-position on dA is very close to its 5'-phosphate (the nitrenoid N and 5'-phosphate O are only 3–4 Å apart in B-form DNA; see Fig. 1) and perhaps also by the length and flexibility of the lysine side chain.

Sequence comparisons (not shown) support the notion that K244 makes a phosphate contact rather than a base contact; although Lys is found at position 244 in only half of the known Rel proteins, the remaining half have another positively charged residue, Arg, in place of Lys. The effects of salt on DNA binding by wild-type and mutant p50 proteins is also consistent with K244 making a phosphate contact. The fact that the wild-type and K244R proteins show similar salt-dependence and overall strength of binding suggests that the most important property of this residue is its positive charge, rather than the specific arrangement of functional groups on the side chain. For Lys residues that make hydrogen-bonding contacts, mutation to Arg can have a dramatically adverse effect on binding; however, Lys and Arg are virtually interchangeable with regard to phosphate contacts (see, for example, [32]). The absence of a positively charged side chain in the K244A and K244M proteins causes a substantial reduction in binding strength at physiological salt concentrations. Interestingly, at low salt concentrations, the wild-type and K244R proteins suffer a loss in DNA affinity, whereas the binding strength of the 244-neutral mutant proteins increases significantly. One possible explanation for this phenomenon might be that position 244 lies nearby other positively charged residues, and the repulsive electrostatic interaction between these residues is poorly screened at low salt. This is not unexpected, since basic residues are often clustered at the protein–DNA interface, especially in the vicinity of the polyanionic phosphodiester backbone [11,22]. Similar effects have been observed elsewhere [33].

All biochemical evidence to date has indicated that the RHR is made up of two subdomains, an amino-terminal segment that contacts DNA directly and a carboxy-terminal segment that is responsible for dimerization [3–10]. Between sequences of the RHR known to be important for DNA binding and dimerization is found a glycine-rich insert (Fig. 5), the sequence of which is not unlike hinge regions commonly found to link domains. This linker might thus reasonably be expected to demarcate the boundary between the DNA-binding and dimerization domains. If this were indeed the case, our results would provide evidence that the dimerization domain of Rel proteins contacts DNA directly. An alternative possibility, which we favor, is that the DNA-binding domain of Rel proteins includes residue 244 and thus extends further toward the carboxy-terminus than previously thought. This view is consistent with truncation studies on p50 suggesting that the amino-terminal boundary of the dimerization domain lies in the region between residues 201 and 253 [9]. Perhaps more compelling is the finding that a p65 fragment corresponding to residues 257–353 of p50 is fully functional for homo- and heterodimerization [10]. Assuming that the boundary of the dimerization and DNA-binding domain lies just past position 244, then the glycine-rich insert must be contained entirely within the DNA-binding domain of p50. Such intra-domain inserts are not uncommon, occurring for example in immunoglobulins, SH3 domains, and protein kinases.

The unusually large size of the Rel homology region has made this family of proteins an attractive and challenging target for studies in protein–DNA interactions. Here we have shown that a novel method of site-specific photocrosslinking can be useful to map proximity relationships in Rel proteins. This method should be applicable not only to other proteins, but to other DNA bases as well.

Significance

Proteins of the Rel transcription factor family are important mediators of cellular defense systems in mammals and are obligate host factors in proliferation of pathogenic viruses. Rel proteins thus hold the key for understanding fundamental aspects of nuclear signalling and are important targets for pharmaceutical intervention in cancer, inflammatory disease and septic shock, among other maladies. In the absence of a high-resolution structure of any member of this family, we and others have turned to chemical and biological approaches toward elucidating structure and function in Rel proteins. In this study, we have developed a photocrosslinking method for the identification of residues that closely approach adenines in DNA, and we have used this method to identify a DNA contact residue of NF-κB p50. This residue, K244, lies substantially farther toward the carboxy-terminus of the protein than any other contact residue discovered to date. This finding suggests that the DNA-binding domain of Rel proteins may be larger than previously thought. By narrowing down more precisely the likely boundary between the DNA-binding and dimerization domains, these studies may enable dissection of the RHR into folded fragments small enough for structural studies by NMR; an improved understanding of the structure of the RHR may eventually enable the design of drugs intended to interfere specifically with the action of Rel family members.

Materials and methods

Materials

All chemicals were purchased from Aldrich with the exception of lithium azide, which was purchased from Kodak (Rochester, NY). Water used in all experiments was doubly distilled and deionized. PAC phosphoramidites were from Pharmacia. Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer. HPLC was performed on a Hewlett-Packard 1090M instrument, which is a generous gift of the manufacturer. The HPLC C-18 reverse phase column was purchased from Beckman (San Ramon, CA). The UV transilluminator (Model 3-3000, 4 x 15 watt 300 nm bulbs, 21 cm x 38 cm UV filter glass) was from Fotodyne (New Berlin, WI). FPLC was performed on a Pharmacia system, using a Pharmacia Mono-Q HR 5/5 column (Piscataway, NJ). Trypsin was from Sigma. The Muta-Gene phagemid kit was from Bio-Rad (Richmond, CA). Autoradiograms were quantified using a Fujix Bio-Imaging Analyzer BAS 2000 from Fuji Medical Systems (Stamford, CT).

Preparation of an N_3dA-containing oligonucleotide

A 23-mer oligonucleotide, 5'-d(AGGGCTGGGG-BrA-TTC-CCCATCTCC)-3', was synthesized as described [25]. An aqueous solution of the BrdA-containing 23-mer (10–20 ml; 20–100 nmol) was transferred to a 5 ml round-bottom flask equipped with a magnetic stirring bar. The contents of the flask were dried *in vacuo* overnight. The flask was released under dry N_2 and equipped with a N_2-filled rubber balloon attached to a ground-glass adapter. All of the remaining steps were carried out in vessels shrouded with aluminum foil. To the flask was added 2.0–4.0 ml of 0.4 M solution of $Et_3NH^+N_3^-$ in DMF, prepared by analogy to a published procedure [34]. The reaction mixture was stirred under N_2 at 75 ± 2 °C for 8 h. We have found it important to avoid the use of rubber septa and especially metal joints and needles. After removal of DMF *in vacuo* at room temperature, the sample was dissolved in 10 ml of 25 mM triethylammonium bicarbonate (TEAB), pH 7.5, and loaded onto a C18 Sep-Pak cartridge (Waters, Milford, MA). The Sep-Pak was washed with 20 ml of 25 mM TEAB, then eluted with 10 ml 30 % (v/v) acetonitrile in 100 mM TEAB. Following lyophilization by SpeedVac (Savant, Farmigdale, NY), the oligonucleotide was dissolved in 100 ml H_2O, and a 1 μl aliquot was used for UV quantification. The sample was analyzed by reverse phase HPLC (gradient from 100 % to 60 % 0.1 M triethylammonium acetate (pH 7) in acetonitrile over 25 min at 1 ml min^{-1}). The oligonucleotide eluted at 10 min, and an impurity derived from reaction of $Et_3NH^+N_3^-$ and DMF eluted at 20 min. The amount of this impurity varies among preparations and is completely absent when the azido transfer reagent is freshly prepared and when the temperature of the reaction is carefully controlled. The impurity does not appear to affect the crosslinking reaction adversely; nonetheless, it can be removed by the HPLC method described above. Typical recovery of the N_3dA-containing 23-mer after HPLC purification is 80–100 %, based on the starting quantity of BrdA-containing 23-mer.

The nucleoside composition of the modified oligonucleotide was examined by HPLC analysis of the products derived from enzymatic digestion [35]. Extinction coefficients for N_3dA and 8-oxo-dA at 254 nm are 3 560 and 7 700, respectively. The identities of the nucleosides were determined by comparison to authentic samples including N_3dA, 8-oxo-dA, BrdA and 8-amino-dA according to their HPLC retention times and photodiode UV spectra; assignments were verified by HPLC co-injection. N_3dA was prepared from BrdA by a procedure similar to that of Buenger and Nair [36]. 8-Amino-dA was prepared from N_3dA by reduction using aqueous DTT [37], and 8-oxo-dA was prepared as described [38].

Analytical-scale photocrosslinking

A fragment of human p50 comprising residues 1–366 expressed in *E. coli* (rp50) [30] was used in all experiments reported here. The N_3dA-containing 23-mer was 5'-end labeled with γ-^{32}P ATP in DTT-free buffer, then annealed to a non-radioactive complementary strand, 5'-d(GGAGATGGGGAATCCCCA-GCCCT)-3', to give the duplex probe containing a single N_3dA residue. For each crosslinking reaction the following were added into an Eppendorf tube (20 μl total volume): 120 fmol of ^{32}P-labeled duplex 32-mer, 5 pmol of rp50 in binding buffer (0.15 mg ml^{-1} BSA, 20 mM HEPES (pH 6.5), 0.2 mM DTT, 9 % (v/v) glycerol, 0.1 mM phenylmethylsulfonyl chloride, 100 mM NaCl, 2.5 mM GTP), and, where applicable, 5 pmol of non-specific or specific competitor DNA. After incubation at room temperature for 30 min, these tubes were uncapped, placed directly under a UV transilluminator, and irradiated at 300 nm for 1 h. SDS-PAGE loading buffer (5 μl) was added, the samples were heated at 95 °C for 15 min, then loaded on a 15 % SDS-PAGE gel. The gel was run at a constant current (20 mA) for 8 h and subjected to quantitative autoradiography using a BioImage analyzer (Fuji).

Large-scale photocrosslinking

The N_3dA-containing 23-mer was annealed to its complement (15 nmol each strand) in 115 μl H_2O. This solution was then added to 100 nmol rp50 in 8 ml binding buffer at 25 °C and

incubated for 0.5 h. The mixture was transferred to a siliconized glass petri dish and placed 5 cm beneath a UV transilluminator (300 nm) and irradiated for 1 h. The reaction mixture was concentrated and washed extensively (~1 000 ml) by ultrafiltration against 3.9 M urea, 20 mM Tris HCl (pH 7.5), 0.5 mM EDTA. The resulting sample (2 ml) was loaded onto an FPLC Mono Q HR 5/5 column, then eluted with a two-step linear gradient from 0 to 0.5 M NaCl (24 ml) followed by 0.5 to 1.0 M NaCl (4 ml) at a rate of 1 ml min^{-1} in urea-free buffer (20 mM Tris HCl (pH 7.5), 0.5 mM EDTA). Fractions from 12–18 min, which contained the covalent rp50–DNA complex, were pooled and precipitated with acetone. The amount of the resulting covalent complex was approximately 1.7 nmol based on UV absorbance. This corresponds to a 9 % yield of crosslinked protein–DNA complex based on the starting duplex 23-mer.

Determination of the crosslinked site

The covalent rp50–DNA complex was dissolved in 50 µl of 0.2 Tris-HCl (pH 9) containing 8 M urea. DTT (5 µl of a 45 mM aqueous solution) and iodoacetamide (5 µl of a 100 mM aqueous solution) were added and the incubation allowed to proceed for 20 min at room temperature. To this mixture, H_2O (126 µl) and trypsin (14 µl of 20 mg ml^{-1} in H_2O) were added and allowed to react for 24 h at 37 °C. This solution was loaded onto FPLC (Mono Q) and eluted with the two-step gradient procedure detailed above. Fractions eluting at 0.4–0.5 M NaCl were pooled, precipitated with ethanol, and subjected to Edman microsequencing as described [39]. The identities, amounts (pmol) of identified amino acids and background (in parentheses) at each cycle were: L, 63.5 (0); E, 25.4 (1.5); P, 26.1 (2.5); V, 34.5 (0.8); V, 59.5 (0.8); S, 13 (1.5); D, 12.2 (1.3); A, 24.1 (2.9); I, 26.1 (0.7); Y, 17.6 (0.6); D, 15.8 (3.5); S, 9.6 (1.6), unidentifiable amino acid, undetermined amount, (0); A, 15.4 (2.9); P, 12.2 (0.9); N, 9.5 (2.3); A, 13.8 (7.4); S, 5.2 (1.4); N, 7.1 (5.4); L, 6.3 (1.3); K, 1.8 (0.5).

Site-directed mutagenesis

Mutations in the rp50 coding sequence were generated using the Bio-Rad Muta-Gene Kit by site-directed mutagenesis on the single-stranded form of the rp50-overproducing phagemid [30].

Electrophoretic mobility shift assays (EMSA)

Proteins (75 nM) were incubated with a ^{32}P end-labeled probe 51-d(AGGGCTGGGGATTCCCCATCTCC)•5'-d(GGAGATGGGGMTCCCCAGCCCT) (25 pM) at room temperature for 30 min in binding buffer (above). NaCl was added in the indicated concentrations. The binding reactions were run alongside one another on a 4 % polyacrylamide gel as described [40]. Protein-bound and unbound DNA were quantified using a Bio-Image analyzer.

Acknowledgements: This work was supported by grants (to G.L.V.) from the American Foundation for AIDS Research, Hoffmann-La Roche (Institute of Chemistry and Medicine), and the National Science Foundation (Presidential Young Investigator Program); we thank Scot Wolfe for assistance with molecular modeling, Chris Larson for discussions and both for a critical reading of the manuscript.

References

1. Sen, R. & Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* **46**, 705–716.
2. Baeuerle, P.A. (1991). The inducible transcription activator NF-κB: regulation by distinct protein subunits. *Biochim. Biophys. Acta* **1072**, 63–80.
3. Kumar, S., Rabson, A.B. & Gélinas, 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.
4. Matthews, J.R., Wakasugi, N., Virelizier, J.-L., Yodoi, J. & Hay, R.T. (1992). Thioredoxin regulates the DNA binding activity of NF-κB by reduction of a disulfide bond involving cysteine 62. *Nucleic Acids Res.* **20**, 3821–3830.
5. Liu, J., Sodeoka, M., Lane, W.S. & Verdine, G.L. (1994). Evidence for a non-α-helical DNA-binding motif in the Rel homology region. *Proc. Natl. Acad. Sci. USA* **91**, 908–912.
6. Coleman, T.A., Kunsch, C., Maher, M., Ruben, S.M. & Rosen, C.A. (1993). Acquisition of NF-κB1-selective DNA binding by substitution of four amino acid residues from NF-κB1 into RelA. *Mol. Cell. Biol.* **13**, 3850–3859.
7. Toledano, M.B., Ghosh, D., Trinh, F. & Leonard, W.J. (1993). N-terminal DNA-binding domains contribute to differential DNA-binding specificities of NB-κB p50 and p65. *Mol. Cell. Biol.* **13**, 852–860.
8. Bressler, P., Brown, K., Timmer, W., Bours, V., Siebenlist, U. & Fauci, A.S. (1993). Mutational analysis of the p50 subunit of NF-κB and inhibition of NF-κB activity by transdominant p50 mutants. *J. Virol.* **67**, 288–293.
9. Logeat, F., *et al.*, & Israel A. (1991). Inhibition of transcription factors belonging to the rel/NF-κB family by a transdominant negative mutant. *EMBO J.* **10**, 1827–1832.
10. Ganchi, P.A., Sun, S.-C., Greene, W.C. & Ballard, D.W. (1993). A novel NF-κB complex containing p65 homodimers: implications for transcriptional control at the level of subunit dimerization. *Mol. Cell. Biol.* **13**, 7826–7835.
11. Pabo, C.O. & Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.
12. Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.
13. Kieran, M., *et al.*, & Israel, A. (1990). The DNA binding subunit of NF-κB is identical to factor κBF1 and homologous to the rel oncogene product. *Cell* **62**, 1007–1018.
14. Allen, T.D., Wick, K.L. & Matthews, K.S. (1991). Identification of amino acids in lac repressor protein cross-linked to operator DNA specifically substituted with bromodeoxyuridine. *J. Biol. Chem.* **266**, 6113–6119.
15. Blatter, E.E., Ebright, Y.W. & Ebright, R.H. (1992). Identification of an amino acid-base contact in the GCN4–DNA complex by bromouracil-mediated photocrosslinking. *Nature* **359**, 650–652.
16. Wick, K.L. & Matthews, K.S. (1991). Interactions between lac repressor protein and site-specific bromodeoxyuridine-substituted operator DNA. *J. Biol. Chem.* **266**, 6106–6112.
17. Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R. & Koch, T.H. (1993). Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* **262**, 1255–1257.
18. Nikiforov, T.T. & Connolly, B.A. (1992). Oligodeoxynucleotides containing 4-thiothymidine and 6-thiodeoxyguanosine as affinity labels for the EcoRV restriction endonuclease and modification methylase. *Nucl. Acids Res.* **20**, 1209–1214.
19. Knowles, J.R. (1972). Photogenerated reagents for biological receptor-site labeling. *Accounts Chem. Res.* **5**, 155–160.
20. Meffert, R., Rathgeber, G., Schafer, H.-J. & Dose, K. (1990). UV-induced cross-linking of Tet repressor to DNA containing tet operator sequences and 8-azidoadenines. *Nucleic Acids Res.* **18**, 6633–6636.
21. Evans, R.K. & Haley, B.E. (1987). Synthesis and biological properties of 5-azido-2'-deoxyuridine 5'-triphosphate, a photoactive nucleotide suitable for making light-sensitive DNA. *Biochemistry* **26**, 269–276.
22. Steitz, T.A. (1990). Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophy.* **23**, 205–280.
23. Gololobov, Y.G., Zhmurova, I.N. & Kasukhin, L.F. (1981). Sixty years of the Staudinger reaction. *Tetrahedron* **37**, 437–472.
24. Liu, J. & Verdine, G.L. (1992). Synthesis of photoactive DNA: incorporation of 8-bromo-2'-deoxyadenosine into synthetic oligodeoxynucleotides. *Tetrahedron Lett.* **33**, 4265–4268.
25. Haley, B. E. (1975). Photoaffinity labeling of adenosine 3',5'-cyclic monophosphate binding sites of human red cell membranes. *Biochemistry* **14**, 3852–3857.
26. Czarnecki, J., Gearlen, R. & Haley, B.E. (1979). Synthesis and use of azido photoaffinity analogs on adenine and guanine. *Methods Enzymol.* **56**, 642–653.
27. Baldwin, A.S., Jr. & Sharp, P.A. (1987). Binding of a nuclear factor to a regulatory sequence in the promoter of the mouse H-2K^b class I major histocompatibility gene. *Mol. Cell. Biol.* **7**, 305–313.
28. Baldwin, A.S., Jr. & 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.

29. 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.
30. Sodeoka, M., Larson, C.J., Chen, L., Lane, W.S. & Verdine, G.L. (1993). Limited proteolysis and site-directed mutagenesis of the NF-κB p50 DNA-binding subunit. *Biomed. Chem. Lett.* **3**, 1095–1100.
31. Bours, V., Villalobos, J., Burd, P.R., Kelly, K. & Siebenlist, U. (1990). Cloning of a mitogen-inducible gene encoding a κB DNA-binding protein with homology to the rel oncogene and to cell-cycle motif. *Nature* **348**, 76–80.
32. Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C. & Pabo, C.O. (1991). The DNA binding arm of repressor: critical contacts from a flexible region. *Science* **254**, 267–270.
33. Jen-Jacobson, L., *et al.*, & Greene P.J. (1983). Coordinate ion pair formation between EcoRI endonuclease and DNA. *J. Biol. Chem.* **258**, 14638–14646.
34. Haley, B.E. (1977). Adenosine 3′,5′-cyclic monophosphate binding sites. *Methods Enzymol.* **46**, 339–346.
35. MacMillan, A.M. & Verdine, G.L. (1990). Synthesis of functionally tethered oligodeoxynucleotides by the convertible nucleoside approach. *J. Org. Chem.* **55**, 5931–5933.
36. Buenger, G.S. & Nair, V. (1990). Dideoxygenated purine nucleosides substituted at the 8-position: chemical synthesis and stability. *Synthesis* 962–966.
37. Cartwright, I.L., Hutchinson, D.W. & Armstrong, V.W. (1976). The reaction between thiols and 8-azidoadenosine derivatives. *Nucleic Acids Res.* **3**, 2331–2339.
38. Guy, A., Duplaa, A.-M., Harel, P. & Téoule, R. (1988). Synthesis and characterization of DNA fragments bearing an adenine radiation product: 7,8-dihydroadenin-8-one. *Helv. Chim. Acta* **71**, 1566–1571.
39. Lane, W.S., Galat, A., Harding, M.W. & Schreiber, S.L. (1991). Complete amino acid sequence of the FK506 and rapamycin binding protein isolated from calf thymus. *J. Protein Chem.* **10**, 151–160.
40. Fried, M. & Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525.

Received: **12 Jul 1994**; revisions requested: **28 Jul 1994**; revisions received: **2 Aug 1994**. Accepted: **2 Aug 1994**.