CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA

Antonella Ronchi, Marianna Bellorini, Nicola Mongelli[1] and Roberto Mantovani*

Dipartimento di Genetica e Biologia dei Microrganismi, Università di Milano, Via Celoria 26, 20133 Milano, Italy and [1]Pharmacia/Farmitalia Carlo Erba, Nerviano Research Center, Via Papa Giovanni XXIII 68, 20014 Nerviano, Italy

ABSTRACT

The CCAAT box is one of the most common promoter elements. The evolutionarily conserved heteromeric factor NF-Y binds this sequence with high affinity and specificity. By comparing the methylation interference patterns of different sites, performing electrophoretic mobility shift assays (EMSA) with IC-substituted oligonucleotides and competition experiments with the minor groove binding (MGB) drugs distamicin A, tallimustine and Hoechst 33258 we show that NF-Y makes key minor groove interactions. Circular permutation assays on four CCAAT boxes, MHC Class II Ea, HSP70, ε-globin and MSV, indicate that NF-Y is able to distort the double helix by angles of 62–82°, depending on the site used, and suggest that nucleotides flanking the CCAAT pentanucleotide influence the degree of bending.

INTRODUCTION

The CCAAT box is a widely distributed regulatory sequence present in several promoters and enhancers. A computer-assisted search on 503 eukaryotic promoters indicated that a high proportion of them (25/30%) harbour this pentanucleotide in the forward or reverse orientation, often at position –60/80 with respect to the start site (1). Unlike other elements composed of palindromic sequences, the CCAAT box does not seem to have a symmetry axis. Several proteins have been reported to bind this or related sequences (2,3). Among these NF-Y, originally identified as a factor recognizing the conserved Y box element in the mouse MHC Class II Ea promoter, has a peculiar and almost absolute requirement for these five nucleotides and a strong preference for additional flanks (2,4). Interestingly, the CCAAT consensus derived by Bucher (PuPuCCAATC/GA/G) fits well with the optimal NF-Y binding site, but not with the consensus of other transcription factors that bind CCAAT-related sequences, such as CTF/NF-1 and C/EBP (1).

NF-Y has been shown to be involved in a number of different systems, in addition to MHC Class II promoters: MSV (2), HSP70 (3), α-, β- and γ-globins (5–7), albumin (8), Thy-1 (9), α-collagen (10), β-actin (11), IL4 (12,13), Gp91phox (14), TSP-1 (15), FGF-4 (16), RSV (17) and TK (18). This CCAAT box binding protein has been identified and, in some cases,

biochemically characterized in several laboratories and, consequently, it has been given different names. Because of the heteromeric nature of the DNA binding complex, of competition analysis with Ea Y box oligonucleotides and supershift experiments with anti-NF-Y monoclonal and polyconal antibodies, CP1 (binding to the MLP, β and γ-globin CCAAT box; 3,7), CBF (to α-collagen; 10), α-CP1 (to α-globin; 5), EFI (to RSV; 17) and CBPtk (to human TK; 18) are different acronyms for the same entity. Therefore it is possible to compare the binding sequences and methylation interference patterns of all these *bona fide* NF-Y binding sites (2,3,5,7,8,11,14–17,19). In all the above-mentioned promoters functional experiments indicate that the CCAAT box and, hence NF-Y, plays an important and sometimes essential role (9,11–16,20–25).

NF-Y is a ubiquitous heteromeric complex formed by at least three subunits, all necessary for DNA binding (3,5,10): NF-YA (HAP2/CBF-B) and NF-YB (HAP3/CBF-A), whose genes have been cloned in a number of species, including maize, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, sea urchin, *Kluyveromyces lactis* (27–34), and the recently discovered HAP5/CBF-C (35,36). Protein alignments of all the subunits evidenced highly conserved domains (70% identity) in all species. Biochemical and genetic analysis dissected the protein–protein and DNA binding subdomains of HAP3 (34) and HAP2/NF-YA (37–39). The remarkable degree of conservation between yeast and man is reminiscent of factors belonging to the basal transcription machinery: TBP for example, the major determinant of Pol II positioning, has been shown to bind DNA via a conserved C-terminal domain that is responsible for making asymmetrical contacts exclusively with the minor groove and for significantly bending the double helix by an angle of 80° (40–43). DNA-dictated asymmetry and protein-mediated distorsion of the DNA are believed to be the rate limiting step in building up a competent transcription complex (44).

The exact mechanisms of transcriptional activation by NF-Y are not known. Although the CCAAT box alone is not able to activate even if multimerized, it appears to increase the activity of neighbouring enhancer motifs (45,46). In the Ea promoter, which has no functional TATA box, NF-Y seems to play an important role in correctly positioning the start site, together with an initiator element (20,47). The strong –60/80 position preference (1) might suggest a structural role in promoter activation.

Because of the evolutionary conservation of the NF-Y subunits across species and the role that the CCAAT sequence plays in

---

Table 1. List of methylation interference data on NF-Y binding sites

	012345678911 01	Reference
Ea	** * TTAACCAATCAGAA AATTGGTTAGTCTT ** °	2
UAS2UP1	** * TCCACCAATCAACG AGGTGGTTAGTTGC ° ** °	19
MLP	° ** * TAAACCAATCACCT ATTTGGTTAGTGGA ** *	3
RSV	** ** TCCACCAATCGGCA AGGTGGTTAGCCGT °° ** *	17
α-GLOBIN	* ** CCAGCCAATGAGTA GGTCGGTTACTCAT ** *	5
γ-GLOBIN	* ** * TTGACCAATAGTCT AACTGGTTATCAGA ** *	6
β-ACTIN	** * GCAGCCAATCAGAG CGTCGGTTAGTCTC ** *	11
ALBUMIN	** **° GGAACCAATGAAAT CCTTGGTTACTTTA ** *	8
TSP1	** ** *** CCGGCCAATGGGCG GGCCGGTTACCCGC ** *	15
Gp91phox	* ** ** TTGACCAATGATTA AACTGGTTACTAAT ** *	14
FGF-4	** ** CCTGCCAATCAGGG GGACGGTTAGTCCC * *	16

diverse promoter contexts in different organisms, we felt that it was important to know more about how NF-Y binds DNA and whether, upon binding, it is able to introduce distortions in the double helix.

MATERIALS AND METHODS

Plasmids

The starting plasmid for the circular permutation assays was pBend2, described by Aryia *et al.* (48). The Ea Y 39mer (2) was inserted by blunt end ligation between the *Sal*I and *Xba*I sites of the polylinker. Oligonucleotides corresponding to the human HSP70 (CCCCAGCCTTCCTTGGACCAATCAGAGGCCA), MSV (TGAACTAACCAATCAGTTCGCT) and human ε-globin (ACACAGGTCAGCCTTGACCAATGACTTTTAAG) CCAAT boxes were blunt end cloned into the *Xba*I site.

Electrophoretic mobility shift assay (EMSA)

Binding reactions for NF-Y were performed according to Chodosh *et al.* (3) and Mantovani *et al.* (39), using recombinant NF-YA (20 ng), purified NF-YB/C (5 ng) and labelled oligonucleotides (10 000 c.p.m.) containing Ea, α-globin and the distal γ-globin (2,7) CCAAT boxes. EMSA with the AP1 site (AGCTT-GATGAGTCAGCCG) were performed by incubating the labelled oligonucleotide in a buffer containing 20% glycerol, 100
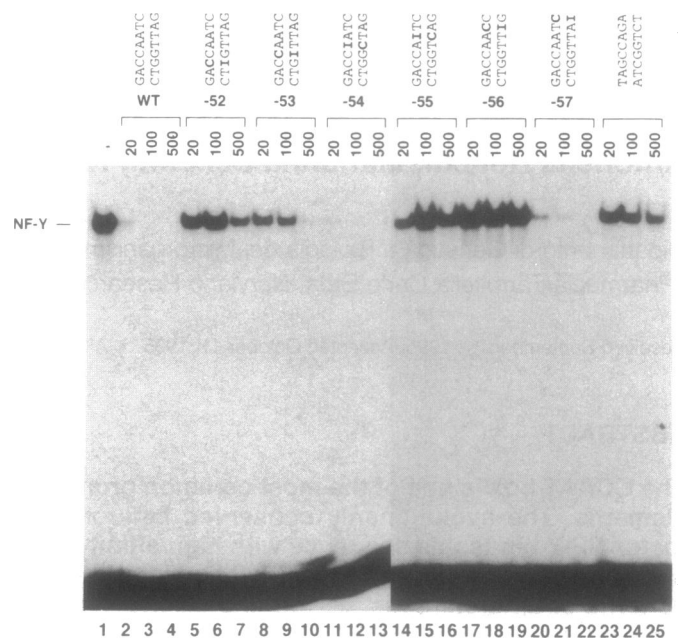

Figure 1. EMSA competition analysis of Y box IC substitution mutants. A labelled Y box 22mer containing the CCAAT box in reverse orientation (2) was used as a probe. Oligonucleotides of identical length, but mutated at a single position, were used as cold competitors. Recombinant NF-YA and purified NF-YB/C were pre-incubated with the indicated molar excess of IC-substituted mutants before addition of the labelled oligonucleotide.

mM KCl, 20 mM HEPES, pH 7.9, 1 mM dTT and 0.5 µl (5 µg) HeLa nuclear extract for 30 min at 25°C.

MGBs were dissolved in DMF (*N,N*-dimethylformamide) at a 10 mM stock concentration (5 mM for tallimustine) and kept at –20°C; fresh dilutions in water were used for EMSA, either before addition or after incubation of the labelled oligonucleotide with the protein (the results were essentially equivalent). For bending assays, NF-Y was incubated with fragments generated by the different cuts in each construct used (pBend2Ea, pBend2ε, pBend2HSP70 and Bend2MSV) and end-labelled with polynucleotide kinase. Acrylamide gels (5%, acrylamide/bis acrylamide ratio 29:1) were run at room temperature. Where indicated an equivalent amount of mutant NF-YA9, described in Mantovani *et al.* (39), was used and also run on 5% gels. Electrophoretic conditions did not significantly modify the results obtained.

Calculations of bending angles

Location of the points of flexure and amplitudes of the bending angles were performed according to the method described in Thompson and Landy (49). Briefly, the mobilities of the NF-Y–DNA complexes were normalized to the mobilities of the corresponding free DNA fragments and bending angles calculated from the ratio between the fastest and the slowest migrating complexes in EMSA and then by linear interpolation between points obtained with A-tract DNA standards (49), according to the formula $\mu_M/\mu_E = \cos \alpha/2$, where μ_M is the relative mobility of the complex exactly in the middle, μ_E is the relative mobility of the complex at the end of the fragment and α is the angle of deviation. To determine bending centres the normalized mobility of each NF-Y–DNA complex was plotted as a function of the
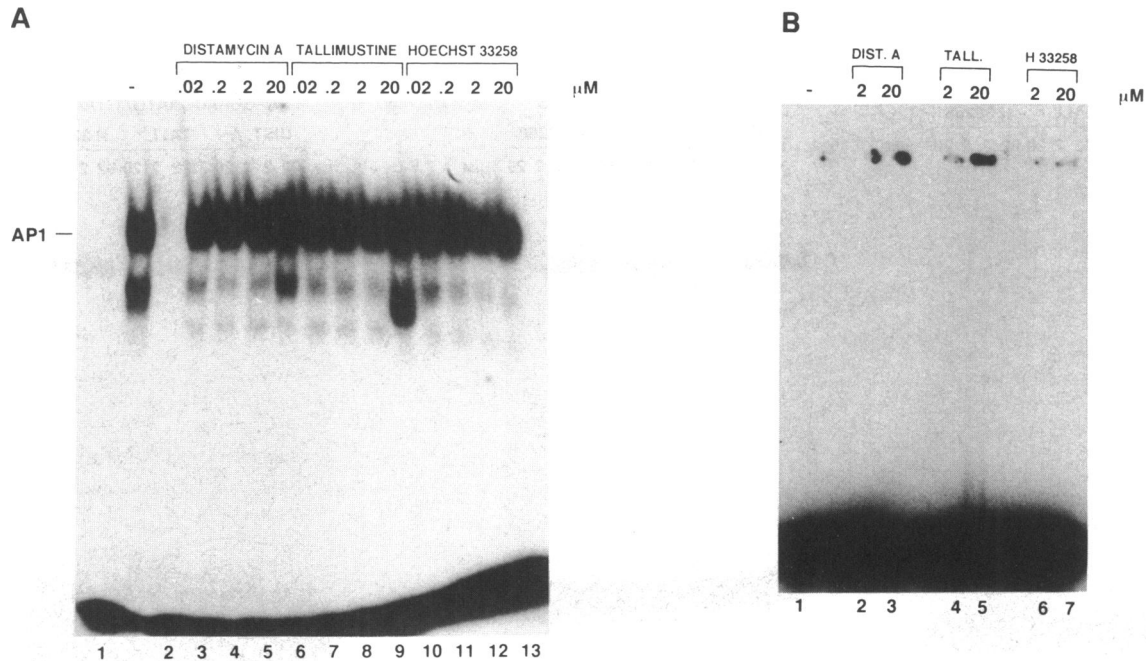
Figure 2. Specificity of MGBs–DNA interactions. (**A**) Lack of EMSA inhibition by MGBs on the AP1 site. A labelled AP1 site corresponding to the collagenase TRE (TPA-responsive element) was incubated with 5 µg HeLa nuclear extract either in the absence (lane 1) or in the presence of increasing concentrations of three MGB drugs: distamycin A (lanes 2–5), tallimustine (lanes 6–9) and Hoechst 33258 (lanes 10–13). (**B**) A labelled Y box oligonucleotide was incubated with the indicated amounts of MGBs in the absence of NF-Y protein.

distance between the centre of the CCAAT sequence and the end of the DNA fragment. The bend was determined as the position at which the NF-Y–DNA complex was at a minimum.

RESULTS

Comparison of the methylation interference patterns of NF-Y on different sites

Because of its wide distribution, high affinity and specificity for CCAAT box sequences, several investigators identified NF-Y in biochemical characterizations of CCAAT box binding proteins involved in the regulation of different promoters. In many such studies methylation interference assays were performed to precisely pinpoint the nucleotides involved in protein–DNA interactions. This assay involves treatment of the target oligonucleotide with dimethylsulfate, which methylates Gs at the N7 position and As at the N3 position, separation of the free from the bound oligonucleotides by EMSA and identification of purine residues necessary for stable binding as missing bands in a sequencing gel. Such experiments not only give precise information about the sequence requirements of the protein, but also suggest whether the protein contacts the major or the minor groove (50): protection of Gs indicates that the protein is contacting the major groove, while a missing A is indicative of minor groove binding.

Table 1 lists the high affinity NF-Y binding sites for which methylation interference data are available. They recognize a single DNA binding factor, variously termed in different laboratories, whose identity with NF-Y has been established by competition with

an Ea Y box oligonucleotide and/or with anti-NF-Y antibodies (5–18). The protected nucleotides in methylation interference patterns are indicated as asterisks and open circles represent partial protection. Several common features are evident.

The two purines at positions +1 and +2 are protected only if one (α- and γ-globin and Gp91) or two Gs (TSP1) are present. Adenines are never protected, whether single (UAS2UP1, RSV, α-globin, γ-globin, β-actin and Gp91phox) or double (Ea, albumin and MLP).

The two Gs on the bottom strand at positions +3/+4 and the two As on the top strand at positions +5/+6 are always protected (with the exception of the +3 G in FGF-4 and the +5 A in Ea). .

The A on the bottom strand at position +7 is completely (MLP, α-globin, γ-globin, β-actin, albumin, TSP1, Gp91 and FGF-4) or partially protected (Ea, UAS2UP1 and RSV).

Protection of Gs on the top strand at positions +8/+10 is evident in some sites (RSV, TSP1, Ea and albumin) but not in others, whereas an A at position +9 is usually protected (Ea, UAS2UP1, β-actin, albumin, Gp91 and MLP), the only exceptions being α-globin and FGF-4. On the bottom strand such positions are never protected, despite the presence of several Gs.

Overall these data suggest that: (i) the NF-Y binding site extends over one turn of the double helix (10–11 bp) and only in MLP, RSV and FGF-4 can partial protections outside the consensus core be scored; (ii) the NF-Y binding site can be tentatively separated into two halves. The left hand PuPuCC (+1/+4) would contact the protein through the major groove, since Gs, but not As, are protected on both the top and bottom strands. On the right hand (+5/+9) NF-Y would make essential contacts in the minor groove
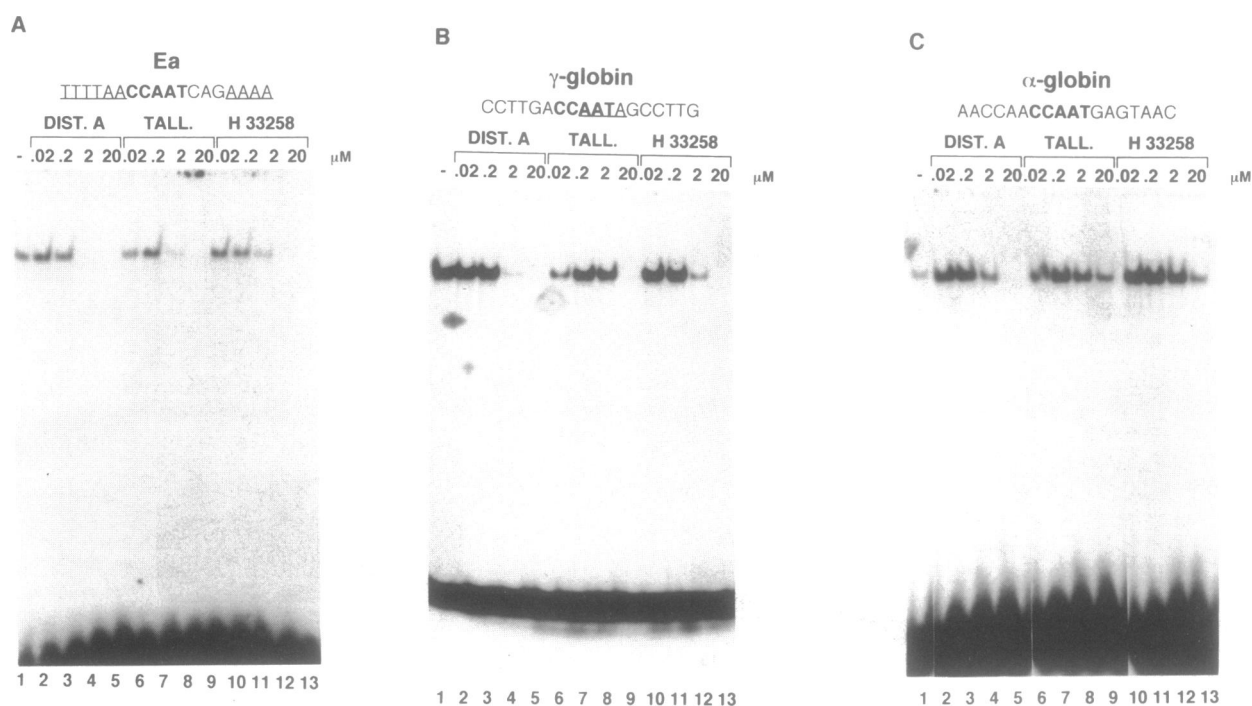
Figure 3. Selective EMSA inhibition by MGBs on NF-Y binding sites. Labelled oligonucleotides corresponding to the Ea Y box (**A**), the γ-globin distal (**B**) and the α-globin CCAAT sequences (**C**) were pre-incubated with the indicated concentrations of distamycin A, tallimustine and Hoechst 33258, before addition of recombinant NF-YA and purified NF-YB/C. AT-rich sequences susceptible to MGB binding are underlined for the three oligonucleotides used.

as well, since As tend to be highly protected both on the top (+5, +6 and +9) and on the bottom strands (+7).

IC substitutions identify essential and non-essential nucleotides

To gain further information on the latter point we exploited the fact that hydrogen bonding on purine and pyrimidine rings in inosine-C pairs resemble G-C in the major groove and A-T in the minor grove (50) and therefore inosine-cytosine mutagenesis can give informations regarding major/minor groove contacts. This feature was successfully used to study the modality of TBP binding to the AdML TATA box (51,52). An oligonucleotide containing IC nucleotide pairs instead of ATs allowed normal interaction with TBP, an indication that only the minor groove was involved in protein–DNA interactions. Such findings were later fully confirmed by crystallographic studies. Similarly, TCF-1 and SRY were also shown to interact mainly with the minor groove (53). It should be noted that the TBP consensus is relatively loose (54–56) and no single nucleotide was shown to be absolutely essential. In the case of NF-Y the sequence-specific requirements are quite high, since saturation mutagenesis studies on a number of different sites clearly established the need for every single base pair of the CCAAT pentanucleotide (2,10,11,17,22) with the exception of the +7 T in two cases, where a C can be found instead (19,57). We introduced IC substitutions into each single nucleotide of the Ea CCAAT box oligonucleotide and performed competition EMSA analysis using a labelled Y box 22mer and purified NF-Y (2,39). Binding to an IC substitution mutant would indicate minor groove binding, while lack of binding would suggest either major groove binding or selective

recognition in the minor groove of a specific determinant in the A/T pair.

Figure 1 shows the results of such competition experiments. Even at a very high molar excess (500-fold) no competition was essentially observed with –52IC, –55IC and –56IC mutant oligonucleotides, while –54IC and –57IC competed essentially as did the wild-type oligonucleotide. An intermediate situation was evident for the –53IC mutation, which competed with a 10-fold lower affinity. Evidence from the methylation interference experiments lends strong support to the idea that the two Cs at the +3/+4 positions are contacted in the major groove, so lack of competition of –52IC and, partially, of –53IC, confirm these data. Normal binding of –54IC strongly points to minor groove contacts, since other mutations at this position alter binding severely (2,17,22). –55IC and –56IC are more puzzling, because if the minor groove is indeed implicated, it has to harbour enough sequence-specific information to account for the almost absolute requirement for these two base pairs. Alternatively, in contradiction of the methylation interference data, this result could indicate that only the major groove is implicated. Finally, good competition of –57IC was not surprising, given the presence in high affinity binding sites of either a C (Ea), an A (γ-globin) or a G (α-globin, albumin).

Effects of minor groove binders on NF-Y binding

To discriminate between the two possibilities emerging from the competition experiments with the –55IC and –56IC substitution mutants we decided to use small drugs known to bind specifically in the minor groove (minor groove binders or MGBs) in EMSA inhibition experiments. Distamycin A is a sequence-specific DNA

binding compound which recognizes AT-rich stretches by interacting non-covalently with the narrower minor groove associated with AT pairs (58). In order to bind DNA it needs only four ATs. The benzoyl nitrogen mustard tallimustine, an alkylating derivative of distamycin A, shows powerful anti-tumor activity and binds similar sequences with a 3-fold lower affinity with respect to distamycin A (59). Hoechst 33258 is another MGB showing preference for AT stretches (60). In control experiments to verify the specificity of such compounds we incubated MGBs at different concentrations with a known major groove binding protein, the Fos/Jun–AP1 complex, and the respective target site (collagenase TRE; see 61). Figure 2A shows that no decrease in the intensity of the AP1 complex is seen, even at high concentrations of the three drugs. The same result was obtained employing the NFkB site (data not shown) and is in line with previous published observations on another DNA binding protein, SP1 (62). In addition, we performed EMSA using the Ea Y box oligonucleotide, which contains two AT-rich stretches, and high concentrations of the three MGBs in the absence of NF-Y. Figure 2B indicates that under the NF-Y binding and electrophoretic conditions used, only tallimustine shows a light smear at 20 μM, an indication of possible dissociation during electrophoresis.

We then felt confident that we could employ these compounds to ascertain whether the AAT right half of the NF-Y site is contacted in the minor groove. For this purpose we decided to use three high affinity sites, Ea and α- and γ-globin, each containing the core CCAAT sequence, but differing considerably in the flanking sequences. While the Ea oligonucleotide contains several binding sites for MGBs outside the CCAAT pentanucleotide, γ-globin, having an A at position +8, only has the AATA stretch (+5/+8) as a possible target for MGB drugs. On the other hand, the α-globin oligonucleotide does not have any run of four consecutive ATs, hence it could not be bound by such compounds.

We used different concentrations of MGBs in EMSA with the above-mentioned oligonucleotides. The results shown in Figure 3A indicate that NF-Y binding to the AT-rich Ea oligonucleotide is completely inhibited at 2 μM distamycin A and at somewhat higher concentrations of tallimustine and Hoechst 33258. Similar data were obtained with the γ-globin oligonucleotide (Fig. 3B); inhibition is quite strong (90%) at 2 μM and complete at 20 μM for the other two MGBs. Finally, the α-globin oligonucleotide shows modest inhibition only at the highest distamycin A concentration and not at all with the other two MGBs (Fig. 3C). The differences between the α- and γ-globin oligonucleotides in the inhibition profiles indicate that NF-Y binding to the AATA sequence, which is the only target for MGBs on the latter oligonucleotide, is strongly diminished when the minor groove is occupied by the drugs. Interestingly, strong inhibition on the Ea oligonucleotide, which can only be bound by MGBs outside the core CCAAT, suggests that minor groove contacts in the flanking sequences might also be important for NF-Y binding, a notion that was previously suggested by some of the mutagenesis studies (2).

NF-Y introduces distortions in the double helix

It is known that several transcription factors introduce flexures in the DNA helix upon binding to their target sites (see 63 for a review). Minor groove binding proteins TBP and HMG box-containing LEF1, HMG1 and SRY introduce large directed distortions (64,65). We therefore wished to verify whether NF-Y could
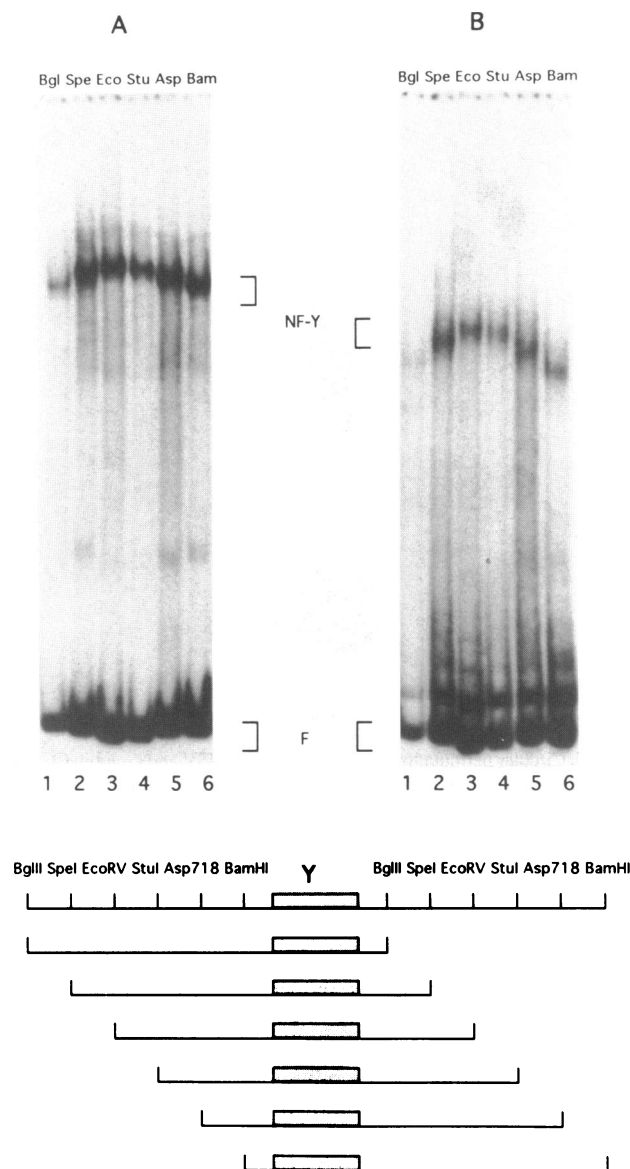

Figure 4. Circular permutation assays of NF-Y. The Y box 39mer was cloned in the bending plasmid pBend2 in the reverse orientation. Six restriction enzymes (indicated in the scheme) were used to generate the 158 bp fragments used in the EMSA. We used recombinant wild-type NF-YA long form (lanes 1–6) and recombinant NF-YA9 mutant, containing only the conserved HAP2 homology region (lanes 7–12), together with purified NF-YB/C.

also bend DNA by performing a circular permutation assay. We cloned the Y box oligonucleotide (2) into the pBend2 plasmid (48) cut with different restriction enzymes, generating fragments of identical length (158 bp) in which the NF-Y binding site is located at different positions with respect to the ends of the fragment, and used them in EMSA with purified NF-YB/C and recombinant NF-YA. Results of such experiments are shown in Figure 4. Fragments in which the Y box is positioned in an eccentric position, *Bam*HI and *Spe*I on the left side and *Asp*718 and *Bgl*II on the right side, show a faster mobility of the NF-Y complex compared with the two central fragments, *Eco*RV and *Stu*1. Using an NF-YA mutant containing only the 57 amino acids sufficient for subunit interaction and DNA binding (YA9; see 39)
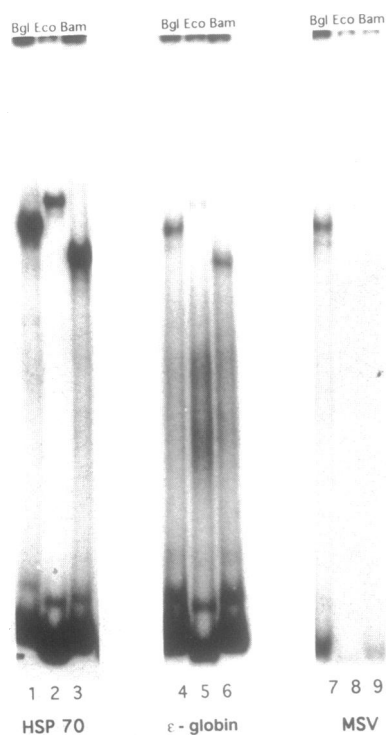
Figure 5. Circular permutation assay with HSP70, MSV and ε-globin CCAAT sites. The *Bam*H1 (lanes 1, 4 and 7), *Eco*RV (lanes 2,5,8) and *Bgl*II (lanes 3,6,9) fragments of pBend2HSP70 (lanes 1–3), pBend2ε (lanes 4–7) and pBend2MSV (lanes 7–9) were incubated with NF-YA9 and NF-YB/C.

Table 2. Mapping of the bending centre of different NF-Y binding sites

very similar results were obtained (Fig. 4B, lanes 1–6). Calculation of the distortion angles induced by NF-Y and NF-YA9 indicated values of 62° and 66° respectively. In both cases the bending centre was close to the core CCAAT sequence, as measured by plotting the mobilities of the NF-Y–DNA complexes as a function of the position of the restriction sites relative to the Y box sequence (Table 2). Incidentally, migration of the free fragments is compatible with a slight degree of intrinsic DNA bending (see the free DNA fragments in Fig. 4).

To further substantiate these results we verified whether other CCAAT boxes embedded in different sequences would also be bent by NF-Y. We employed sites that have been shown to have different affinities for NF-Y: the human HSP70 (high affinity; R.Mantovani, unpublished), the MSV (medium affinity; 2) and the ε-globin (low affinity; A.Ronchi and R.Mantovani, in preparation) sites. As shown in Figure 5, NF-Y clearly distorts the DNA molecules, by 80° and 82° respectively for the wild-type NF-Y and NF-YA9 proteins used, similar among them, but slightly different from the values obtained with the Ea Y box. In all these experiments the flexure centre is positioned in the CCAAT pentanucleotide (Table 2). The sharper angles observed with these probes (20° difference) are most likely due to sequences surrounding the conserved AACCAAT sequence.

Taken together, these results indicate that NF-Y is able to significantly bend DNA upon binding and that the Q-rich and S/T-rich regions of NF-YA are largely expendable. The centre of distortion is located in the conserved CCAAT region (see Table 2) and the nature of the neighbouring sequences remarkably influences the degree of bending.
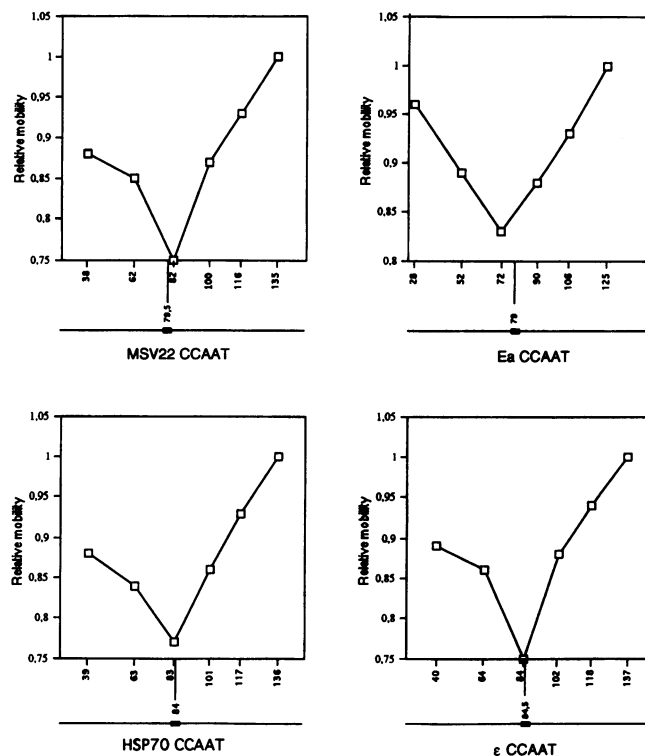
DISCUSSION

In this report we have investigated some of the *in vitro* binding features of the CCAAT box binding protein NF-Y and reached a few conclusions that can be summarized in the following points.

NF-Y binds both in the major and minor grooves of the double helix, based on the methylation interference patterns of several *bona fide* NF-Y binding sites, on competition experiments with three MGBs and, partially, on IC substitution mutagenesis. MGB experiments with the Ea and γ-globin CCAAT boxes show that inhibitory concentrations are similar to those required to prevent interaction of TBP with the minor groove of TATA sequences (62,66). This assay appears to be rather specific, since proteins that recognize the major groove, such as Fos/Jun family members or SP1, are not inhibited by these compounds (62,67). Crystallographic studies on several sequence-specific DNA binding proteins interacting with their target sequence determined that contacts are made only in the major groove, where most of the determinants allowing proteins to discriminate between the different nucleotides are located (68). Proteins that bind DNA as dimers, such as members of the leucine zipper and HLH families, are included in the list. However, notable exceptions are homeodomain (HD) proteins, recognizing both the major and, through the N-terminal part, the minor groove, HMG proteins and TBP, which exclusively makes sequence-specific contacts in the minor groove (40–43).

It has recently been shown that MGBs can inhibit TBP binding and basal *in vitro* transcription (62). Here we show that NF-Y binding can also be prone to MGB interference at similar concentrations, but only at selected sites, indicating that even within a single DNA binding protein inhibitory doses can vary

over two orders of magnitude and implying a remarkable degree of selectivity for these drugs. It will be interesting to test the role of such compounds in functional CCAAT-dependant assays.

IC substitution experiments prove that despite minor groove binding, the +6 A and +7 T retain remarkable sequence specificity. Other minor groove binding factors, like TBP and HMG, are far less susceptible to single nucleotide changes (53–56), in fact, it has been shown that the entire TATA box can be mutated into IC sequences and still bind TBP normally (51,52). This is certainly not possible for NF-Y, which has a high degree of discriminatory power not only in the major groove, but for minor groove constituents as well.

Both methylation interference and competition with MGBs suggest that the NF-Y binding site can be divided in two halves: the 5′-part (PuPuCC) is in contact with the major groove, while the 3′-part (AATC/AG/G) interacts with the minor groove. NF-Y seems to lie on one side, contacting at least one full turn of the double helix. The apparent asymmetry of the CCAAT box probably mirrors asymmetric contact points on the NF-Y heteromeric protein complex. Perhaps the best known example of an asymmetric binding site is represented by the TATA box: TBP however, thanks to two direct repeats present in the C-terminal conserved domain, contains a remarkable symmetry in the protein surfaces presented for association with other polypeptides (40–43). The polarity comes from the non-conserved N-terminal region. No repeat is found in the DNA binding domains of the three NF-Y subunits contacting DNA. We can conclude that asymmetry is the rule for this transcription factor, both at the DNA and at the protein levels. This finding could imply important functional consequences. The CCAAT box, in fact, is present in both the forward (globins, albumin and β-actin) and reverse orientations (MHC Class II), statistically equally well represented (1). Moreover, the distance between this element and the binding sites of several other factors in different promoters is important, including in MHC Class II promoters. In some cases NF-Y seems to favour the binding of transcription factors to nearby sequences: RFX to the Dra X box (69) and C-EBP to the albumin D box (70). The polarity of NF-Y binding could then be essential for protein–protein interactions and influence the general promoter architecture. This might indicate that different parts of the complex are specifically able to contact different classes of activators.

As a result of NF-Y binding the DNA is distorted to a degree that is somewhat dependant on the composition of neighbouring sequences, varying between 62° and 82°. The centre of the flexure is the CCAAT box, itself a sequence that seems to be to some extent intrinsically bent. The bending capacity is shared by a number of other transcription factors and seems to be a common theme among proteins that do not have strong activating surfaces (40–43,53,64,65). The degree of bending, which is relatively large compared with other transcription factors, such as ER, Ig/IBP, LAP and OCT1, all inducing angles between 30° and 50°, resembles more the sharper angles generated by TBP (80°; 40–43), YY1 (78°; 71) and HMG (75/130°; 64,65). Such proteins are known to play a general role in the organization of the three-dimensional promoter structure, rather than directly activating transcription (40–43,64,66,71). NF-Y is not able to activate alone, but rather needs other activators nearby (45,46). It is conceivable that this cooperative function is in part elicited by indirect facilitation of protein–protein interactions, brought about by NF-Y-induced DNA distortions at the CCAAT box. Moreover,

several promoters, including γ-globin, gp91phox, IL4, DNA Pol α and TK (7,13,14,18), contain multiple NF-Y binding sites and therefore are probably prone to multiple NF-Y-dependent levels of organization.

The Q-rich and S/T-rich domains of NF-YA are expendable for DNA bending and do not seem to significantly influence the degree of distortion. The phylogenetically conserved subunit interaction and DNA binding regions of the NF-Y components are necessary and sufficient, a feature found in some, but not all DNA binding proteins: Fos/Jun and other members of the leucine zipper family, for example, show a clear influence of other protein domains on the degree of DNA flexure (72,73).

If protein sequences outside the conserved HAP homology domains are neutral, DNA sequences flanking the CCAAT box appear to play a role in determining the amplitude of the angle. The Ea Y box sequences flanked by AT-rich runs reduce the angle, possibly by generating intrinsic bends at both the 5′- and 3′-ends of the pentanucleotide. Indeed, we note that such bends are in agreement with the rules determined by the studies of Koo *et al.* (74). On the other hand, when embedded in GC-rich stiffening sequences NF-Y induces wider angles. Surrounding nucleotides, without making direct contacts with the protein, are probably far from being neutral in terms of the overall effect of NF-Y binding on promoter conformation. It is known that in all MHC Class II genes the distance between the X and the Y box is invariant in length, but variable in sequence (75). We note that AT-rich stretches are conserved at both the 5′-end, between X and Y, and the 3′-end of the Y box, suggesting that they are important for the correct spatial relationships between NF-Y and X box binding activators on one side (69) and, possibly, between NF-Y and general transcription factors responsible for Pol II positioning on the other.

ACKNOWLEDGEMENTS

REFERENCES

1 Bucher,P. (1990) *J. Mol. Biol.*, **212**, 563–578.
2 Dorn,A., Bollekens,J., Staub,A., Benoist,C. and Mathis,D. (1987) *Cell*, **50**, 863–872.
3 Chodosh,L.A., Baldwin,A.S., Carthew,R.W. and Sharp,P. (1988) *Cell*, **53**, 11–24.
4 Hooft van Huijsduijnen,R., Bollekens,J., Dorn,A., Benoist,C. and Mathis,D. *Nucleic Acids Res.*, **15**, 7265–7272.
5 Kim,C.G. and Sheffery,M. (1990) *J. Biol. Chem.*, **265**, 13362–13369.
6 de Boer,E., Antoniou,M., Mignotte,V., Wall,L. and Grosveld,F. (1988) *EMBO J.*, **7**, 4203–4212.
7 Superti-Furga,G., Barberis,A., Schaffner,G. and Busslinger,M. (1988) *EMBO J.*, **7**, 3099–3107.
8 Raymondjean,M., Cereghini,S. and Yaniv,M. (1988) *Proc. Natl Acad. Sci USA*, **85**, 757–761.
9 Spanopoulou,E., Giguere,V. and Grosveld,F. (1991) *Mol. Cell. Biol.*, **11**, 2216–2228.
10 Hatamochi,A., Golumbek,P., Van Schaftingen,E. and de Crombrugghe,B. (1987) *J. Biol. Chem.*, **263**, 5940–5947.

11 Danilition,S.L., Frederickson,R.M., Taylor,C.Y. and Miyamoto,N.G. (1991) *Nucleic Acids Res.*, **19**, 6913–6922.

12 Szabo,S.J., Gold,J.,Murphy,T.L. and Murphy,K.M. (1993) *Mol. Cell. Biol.*, **13**, 4793–4805.

13 Weber,M.L., Davydov,I.V., Krafft,H. and Krammer,P.H. (1994) *J. Immunol.*, **153**, 4122–4133.

14 Skalnik,D.G., Strauss,E.C. and Orkin,S.H. (1991) *J. Biol. Chem.*, **266**, 16736–16744.

15 Framson,P. and Bornstein,P. (1993) *J. Biol Chem.*, **268**, 4989–4996.

16 Hasan,S., Koda,T. and Kakinuma,M. (1994) *J. Biol. Chem.*, **269**, 25042–25048.

17 Faber,M. and Sealy,L. (1990) *J. Biol. Chem.*, **265**, 22243–22254.

18 Chang,Z.F. and Liu,C.-J. (1994) *J. Biol. Chem.*, **269**, 17893–17898.

19 Olesen,J., Hahn,S. and Guarente,L. (1987) *Cell*, **51**, 953–961.

20 Dorn,A., Durand,B., Marfing,C., LeMeur,M., Benoist,C. and Mathis,D. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6249–6253.

21 Karsenty,G., Golumbek,P. and deCrombrugghe,B. (1988) *J. Biol. Chem.*, **263**, 13909–13915.

22 Wuarin,J., Muller,C. and Schibler,U. (1990) *J. Mol. Biol.*, **214**, 865–874.

23 Tronche,F., Rollier,A., Sourdive,D., Cereghini,S. and Yaniv,M. (1991) *J. Mol. Biol.*, **221**, 31–43.

24 Myers,R., Tilly,K. and Maniatis,T. (1986) *Science*, **232**, 613–618.

25 Greene,J.M., Larin,Z., Taylor,I.C.A., Prentice,H., Gwinn,K.A. and Kingston,R. (1987) *Mol. Cell. Biol.*, **7**, 3646–3655.

26 Hooft van Huijsduijnen,R., Li,X.Y., Black,D., Matthes,H., Benoist,C. and Mathis,D. (1990) *EMBO J.*, **9**, 3119–3127.

27 Maity,S., Vuorio,T. and de Crombrugghe,B. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5378–5382.

28 Vuorio,T., Maity,S. and de Crombrugghe,B. (1990) *J. Biol. Chem.*, **265**, 22480–22486.

29 Li,X.-Y., Mantovani,R., Hooft van Huijsduijnen,R., Andre,I., Benoist,C. and Mathis,D. (1992) *Nucleic Acids Res.*, **20**, 1087–1091.

30 Pinkham,J.L., Olesen,S., Guarente,L. (1987) *Mol. Cell. Biol.*, **7**, 578–585.

31 Hahn,S., Pinkham,J., Weil,R., Miller,R., Guarente,L. (1988) *Mol. Cell. Biol.*, **8**, 655–663.

32 Olesen,S., Fikes,J.D., Guarente,L. (1991) *Mol. Cell. Biol.*, **11**, 611–619.

33 Li,X., Kalasapudis,S. and Childs,J. (1993) *Nucleic Acids Res.*, **21**, 4639.

34 Xing,Y., Fikes,J.D. and Guarente,L. (1993) *EMBO J.*, **12**, 4647–4655.

35 McNabb,D.S., Xing,Y. and Guarente,L. (1995) *Genes Dev.*, **9**, 47–58.

36 Sinha,S., Maity,S.N., Lu,J. and de Crombrugghe,B. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 1624–1628.

37 Xing,Y., Zhang,S., Olesen,J., Rich,A. and Guarente,J. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3009–3013.

38 Maity,S. and de Crombrugge,B. (1992) *J. Biol. Chem.*, **267**, 8286–8292.

39 Mantovani,R., Li,X.-Y, Pessara,U., Hooft van Huijsduijnen,R., Benoist,C. and Mathis,D. (1994) *J. Biol. Chem.*, **269**, 20340–20346.

40 Nikolov,D.B., Hu,S.-H., Lin,J., Gash,A., Hoffman,A., Horikoshi, M., Chua,N.-H., Roeder,R.G. and Burley,S.K. (1992) *Nature*, **360**, 40–46.

41 Chasman,D.I., Flaherty,K.M., Sharp,P. and Kornberg,R.D. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8174–8178.

42 Kim,Y., Geiger,J.H., Hahn,S., Sigler,P. (1993) *Nature*, **365**, 512–520.

43 Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) *Nature*, **365**, 520–527.

44 Zawel,L., and Reinberg,D. (1993) *Prog. Nucleic Acid Res.*, **44**, 67–108.

45 Strahle,U., Schmid,W. and Schutz,G. (1988) *EMBO J.*, **11**, 3389–3395.

46 Wang,W. and Gralla,J. (1991) *Mol. Cell. Biol.*, **11**, 4561–4571.

47 Mantovani,R., Tora,L., Moncollin,V., Egly,J.M., Benoist,C. and Mathis,D. (1993) *Nucleic Acids Res.*, **21**, 4873–4878.

48 Kim,J., Zwieb,C., Wu,C. and Arya,S. (1989) *Gene*, **85**, 15–23.

49 Thompson,J.F. and Landy,A. (1988) *Nucleic Acids Res.*, **16**, 9687–9705.

50 Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 122–126.

51 Starr,D.B and Hawley,D.K. (1991) *Cell*, **67**, 1231–1240.

52 Lee,D.K., Horikoshi,M. and Roeder,R.G. (1991) *Cell*, **67**, 1241–1250.

53 van de Wetering,M. and Clevers,H. (1992) *EMBO J.*, **11**, 3039–3044.

54 Hahn,S., Buratowski,S., Sharp,P. and Guarente,L. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5718–5722.

55 Wobbe,C.R., and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 3859–3867.

56 Wong,J.M. and Bateman,E. (1994) *Nucleic Acids Res.*, **22**, 1890–1896.

57 Gu,Z., Plaza,S., Perros,M., Cziepluch,C., Rommelaere,J. and Cornelis,J. (1995) *J. Virol.*, **69**, 239–246.

58 Coll,M., Frederick,C.A., Wang,A.H.-J. and Rich,A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.

59 Broggini,M., Coley,H., Mongelli,N., Grandi,G., Wyatt,M.D., Hartley,J.A. and D'Incalci,M. (1995) *Nucleic Acids Res.*, **23**, 81–87.

60 Churchill,M. and Travers,A. (1991) *Trends Biochem. Sci.*, **16**, 92–97.

61 Risse,G., Jooss,K., Neuberg,M., Brullel,H.J. and Muller,R. (1989) *EMBO J.*, **8**, 3825–3832.

62 Bellorini,M., Moncollin,V., Mongelli,N., D'Incalci,M. and Mantovani,R. (1995) *Nucleic Acids Res.*, **23**, 1657–1663.

63 Perez-Martin,J., Rojo,F. and deLorenzo,V. (1994) *Microbiol. Rev.*, **58**, 269–290.

64 Giese,K., Cox,J. and Grossschedl,R. (1992) *Cell*, **69**, 185–195.

65 Ferrari,S., Harley,V.R., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M. (1992) *EMBO J.*, **11**, 4497–4506.

66 Broggini,M., Ponti,M., Ottolenghi,S., D'Incalci,M., Mongelli,N. and Mantovani,R. (1989) *Nucleic Acids Res.*, **17**, 1051–1059.

67 Chiang,S.-Y., Welch,J., Raucher,F.J. and Beerman,T.A. (1994) *Biochemistry*, **33**, 7033–7040.

68 Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.

69 Reith,M., and Mach,B. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 554–558.

70 Milos,P. and Zaret,K.S. (1992) *Genes Dev.*, **6**, 991–1004.

71 Natesan,S. and Gilman,M.Z. (1993) *Genes Dev.*, **7**, 2497–2509.

72 Kerppola,T.K. and Curran,T. (1991) *Cell*, **66**, 327–336.

73 Kerppola,T.K. and Curran,T. (1993) *Mol. Cell Biol.*, **13**, 5479–5489.

74 Koo,H-S., Wu,H.-M. and Crothers,D.M. (1986) *Nature*, **320**, 501–506.

75 Benoist,C. and Mathis,D. (1990) *Annu. Rev. Immunol.*, **8**, 681–715.