Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 3992–3996, May 1990
Biochemistry

Alkylation interference identifies essential DNA contacts for sequence-specific binding of the eukaryotic transcription factor C/EBP

(DNA–protein interaction/"leucine zipper" protein/ethylation)

JULIE A. NYE AND BARBARA J. GRAVES

Department of Cellular, Viral and Molecular Biology, University of Utah School of Medicine, Salt Lake City, UT 84132

Communicated by Brian W. Matthews, March 8, 1990

ABSTRACT **Transcriptional regulators containing a "leucine zipper" and a flanking basic domain belong to a recently identified class of DNA-binding proteins. We have mapped the essential DNA contacts of one member of this group, the eukaryotic transcription factor C/EBP. Methylation and ethylation interference experiments detected major groove contacts over a full turn of the DNA helix on both an asymmetric and a symmetric C/EBP binding site. The contacts essential for C/EBP binding have two-fold symmetry yet differ significantly from the contacts of other dimeric DNA-binding proteins, including those bearing helix–turn–helix motifs and the type I restriction endonuclease *Eco*RI.**

The eukaryotic transcription factors Jun, Fos, GCN4, and C/EBP belong to a group of DNA-binding proteins that contain a "leucine zipper" and a flanking basic domain (1). Both components are required for DNA binding. The leucine zipper forms a coiled coil of α-helices (2, 3) and mediates requisite dimerization (4, 5). The basic region contributes to DNA recognition (4, 6), but the details of its structure and its contacts with DNA are unknown. The mechanism of specific DNA recognition is likely to be different from that of the helix–turn–helix or zinc-finger motifs because no similarity in protein sequence is observed (1, 7).

To characterize the sequence-specific recognition of DNA by this class of protein, we have used both guanine-methylation and phosphate-ethylation interference (8) to map the essential contacts of C/EBP on the DNA helix. C/EBP was initially characterized as a rat liver nuclear protein that recognizes both the CAAT box and the enhancer core of several viral transcriptional control elements (9–11). Our study focuses on two CAAT-box sites that share the pentanucleotide 5'-GCAAT-3'. One, within the mouse transferrin (mTf) promoter, displays almost perfect dyad symmetry (12) and another, within the promoter of the Moloney murine sarcoma virus (MSV) long terminal repeat (LTR) (10), shows no striking symmetry.

In this paper, we demonstrate that the essential DNA contacts made by C/EBP display two-fold symmetry and include major groove contacts over 12–14 contiguous nucleotide pairs. A similar set of essential DNA contacts is observed on both a symmetrical and an asymmetrical C/EBP binding site. A comparison of the data to previously described DNA–protein interfaces indicates that the leucine zipper/basic region proteins bind DNA in a structurally distinct manner. The implications for sequence-specific DNA recognition by C/EBP are discussed.

MATERIALS AND METHODS

Plasmids and DNA Molecules. A *Hin*dIII/*Sst* I restriction fragment bearing the MSV LTR promoter sequences between −110 and −31 (10) was subcloned into pEMBL 19. The 85-base-pair (bp) *Hin*dIII/*Eco*RI restriction fragment was radiolabeled at either the *Hin*dIII site on the top strand or at the pEMBL *Eco*RI site on the bottom strand with T4 polynucleotide kinase and [γ-^{32}P]ATP. The promoter of the mTf gene, subcloned in pUC13, was obtained from G. Stanley McKnight (12). A *Hin*fI/*Hae* III restriction fragment, bearing promoter sequences −180 to −44, was radiolabeled at the *Hin*fI site for top-strand analysis. The same site was radiolabeled with *Escherichia coli* DNA polymerase (Klenow fragment) and [α-^{32}P]dATP for bottom-strand analysis.

Alkylation Interference. End-labeled fragments (0.5–1.0 pmol) were methylated with dimethyl sulfate (Aldrich) or ethylated with *N*-ethyl-*N*-nitrosourea (Sigma) to yield less than one modification per molecule (8, 13). C/EBP was produced in *E. coli* from the full-length rat cDNA clone (4) by using the pET expression system (14). The 42-kDa C/EBP polypeptide constituted less than 5% of the total protein in bacterial cell lysates. Extracts were minimally fractionated by DEAE-cellulose chromatography as described (1). C/EBP was bound to alkylated DNA by combining DNA (15 fmol; 2.5×10^4 cpm) with 0.3–1.6 μg of bacterial extract in 50 μl of binding buffer [10 mM Tris·HCl (pH 7.9), 5 mM $MgCl_2$, 74 mM KCl, 1 mM dithiothreitol, 0.5 mM EDTA, 10% (vol/vol) glycerol, and poly(dI-dC) (100 μg/ml)]. Extract quantity was chosen to achieve 10–30% DNA occupancy at equilibrium. LTR promoter binding reactions contained 5-fold more extract than mTf promoter reactions since the mTf site binds C/EBP ≈5-fold more tightly than the LTR site. Samples were incubated for 30 min at 4°C and then electrophoresed on 5% (30:0.8) native acrylamide gels using a 45 mM Tris borate (pH 8.3) buffer (15). Unbound and protein-bound DNA fractions were identified by their different mobilities and were extracted from the gel. Strands were cleaved at phosphotriesters by NaOH and at methylated guanine residues by piperidine (Fisher) (8, 13). Samples of bound and unbound DNA containing equivalent amounts of radioactivity were electrophoresed on denaturing 9% acrylamide gels. Gels were exposed to preflashed Cronex film (DuPont) with intensifying screens. Autoradiographs were scanned with a Bio-Rad densitometer, and film densities were plotted with IGOR software (WaveMetrics, Lake Oswego, OR). Peak heights on densitometric tracings were used for quantitation.

RESULTS

Fig. 1 shows the sequences of the C/EBP binding sites in the mTf promoter and the MSV LTR promoter. The conserved 5'-GCAAT-3' motifs are indicated by the arrows. In the mTf promoter, the bottom strand contains two 5'-CCAAT-3' pentanucleotides. The one located between −94 to −98 is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LTR, long terminal repeat; MSV, murine sarcoma virus; mTf, mouse transferrin.

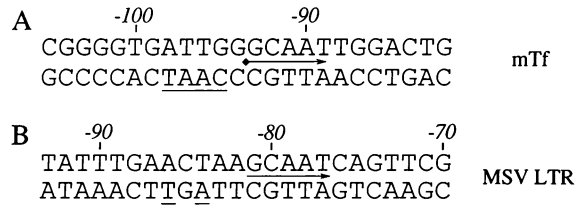

FIG. 1. C/EBP binding sites. (*A*) mTf promoter. (*B*) MSV LTR promoter. Coordinates are the promoter positions relative to the start site of transcription. Bacterially expressed C/EBP protects these promoter regions from DNase I digestion (data not shown). Sites are aligned by the conserved GCAAT motif, indicated by the arrows. Positions that show dyad symmetry with the GCAAT motif are underlined. ◆, Proposed dyad axis of the mTf site. The wild-type LTR sequence (−81 to −76) contains a 5′-CCAAT-3′ pentanucleotide. In this study, we used a mutant LTR that has a transversion of the first cytosine and displays a 10-fold higher affinity for C/EBP than the wild-type LTR (10).

oriented such that the site displays almost perfect dyad symmetry over 10 bp (Fig. 1*A*). In the promoter of the MSV LTR, only two matches to the GCAAT motif are symmetrically displayed (Fig. 1*B*). We term the LTR site asymmetric and the mTf site symmetric.

Restriction fragments bearing either the mTf or the LTR site were end-labeled and sparingly modified with either dimethyl sulfate or *N*-ethyl-*N*-nitrosourea (8, 13). The population of modified DNA molecules was then incubated with C/EBP. Protein-bound DNA was separated from unbound DNA on low-ionic-strength polyacrylamide gels (15). Modified sites were cleaved and mapped with respect to the labeled end by electrophoresis on denaturing polyacrylamide gels. Modifications that reduce the affinity of the binding site for C/EBP are enriched in the unbound DNA and depleted in the bound DNA fractions.

Alkylation Interference on the Asymmetric Site. Alkylation interference data on the LTR site are shown in Fig. 2 *A* and *B*. Interference is detected by the higher intensity of a band in the unbound lane relative to the intensity of a band in the bound lane. Methylation of the guanine at position −81 on the top strand in the LTR promoter interferes with binding, whereas modification of the guanine residue at position −74 or −88 has no effect on C/EBP binding (Fig. 2*A*). On the LTR bottom strand, interference is observed at guanine residues at positions −80 and −76 (Fig. 2*B*). Methylation of flanking guanine residues at positions −71 and −85 does not affect C/EBP binding. Significant interference with C/EBP binding (greater than 2-fold effect) results from ethylation of nine successive phosphates on both the top and bottom strands of the LTR site. Fig. 3 *A* and *B* displays densitometric tracings of ethylation interference data for the LTR site. The most severe interference maps near the GCAAT pentanucleotide on both strands.

Alkylation Interference on the Symmetric Site. Methylation and ethylation interference data for the more symmetrical mTf promoter site present a complementary picture (Fig. 2 *C* and *D*). Methylation of the four guanine residues near the center of the dyad axis (−93 to −95 on the top strand, −92 on the bottom strand) interferes with C/EBP binding. Modification of an upstream guanine residue on the top strand (−99) also disrupts binding, but methylation of guanine residues either further upstream or downstream has no detectable effect in this assay (Fig. 2 *C* and *D*). In addition, ethylation of 12 different phosphates on the top strand and 11 on the bottom strand significantly interferes with C/EBP binding (Fig. 2 *C* and *D*). A quantitative analysis of the mTf promoter ethylation interference data is presented in Fig. 3 *C* and *D*. Severely interfering ethylations map to both halves of this symmetrical site.

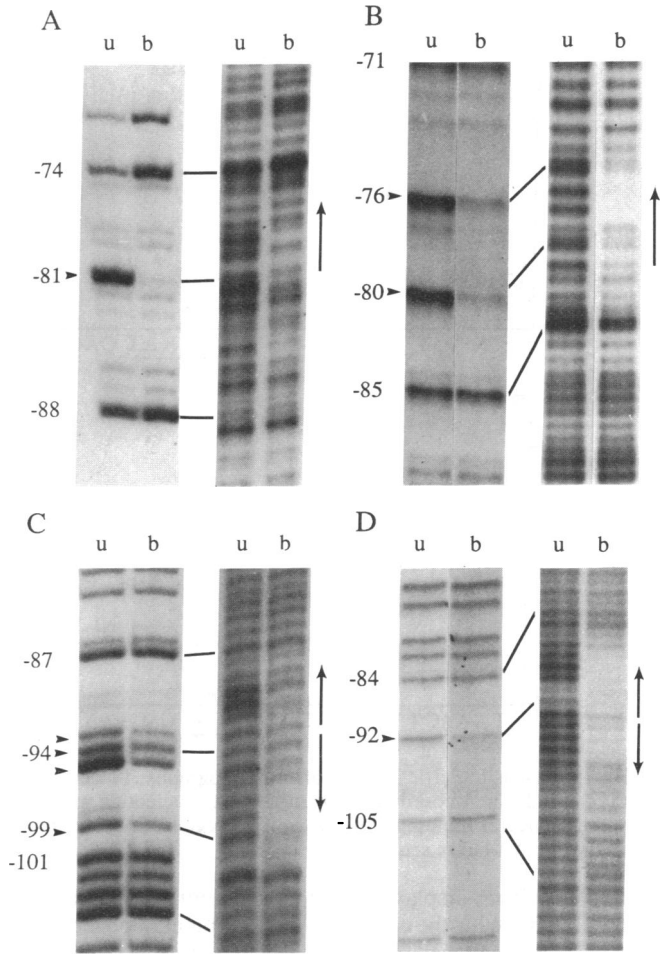

FIG. 2. Alkylation interference analysis. The top (*A*) and bottom (*B*) strands of the LTR site and the top (*C*) and bottom (*D*) strands of the mTf site were subjected to methylation (*Left*) and ethylation (*Right*) interference. Coordinates are the position of promoter nucleotides numbered relative to the start site of transcription as determined by coelectrophoresis with chemically sequenced promoter fragments. Lane u, DNA (unbound) that was incubated with protein but had the mobility of free DNA on nondenaturing gels. Lane b, DNA (bound) recovered from slower migrating DNA–protein complexes. Vertical arrows, conserved CAAT-box sequences; horizonal arrowheads, methylated guanine residues that interfere with binding. The weaker intensity of bands in the unbound methylation lane in *A* is due to underloading of this lane. In general, ethylated DNA lanes have diffuse bands due to breakage of the DNA backbone on either side of the ethylated phosphate. *N*-Ethyl-*N*-nitrosourea also can react with O^6 of guanine (16); therefore, guanine positions are sometimes overrepresented in the ethylated DNA ladder.

Two-Fold Symmetry of Contacts. Fig. 4 summarizes the quantitative analysis of ethylation interference effects on the two binding sites. In each promoter, the pattern of proposed phosphate contacts has two-fold symmetry, with the dyad axis positioned 5′ to the GCAAT pentanucleotide. Note that the distribution of the most severely interfering ethylations highlights the two-fold symmetry. These important phosphate contacts map in each of the four half-sites to similar positions with respect to the dyad axis (positions 2 and 3 on the 5′-GCAAT-3′ sequence or its equivalent; positions 4 and 5 on the 3′-CGTTA-5′ sequence or its equivalent). These data indicate that C/EBP is similarly disposed on the two binding sites. We used this similarity to assign coordinates for further discussion of the two sites. Nucleotides of the GCAAT half-sites are designated 1R–8R and those of the non-GCAAT half-sites are designated 8L–1L. A striking finding is that the

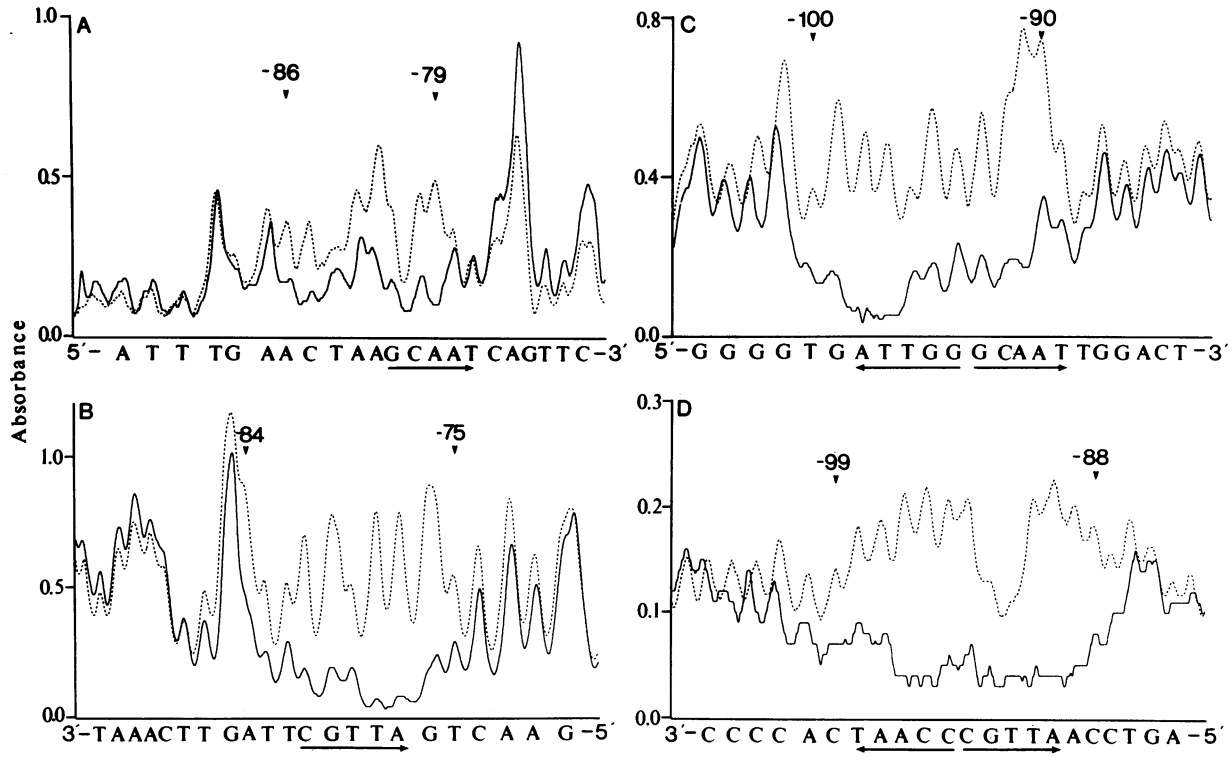

FIG. 3. Densitometric analysis of autoradiographs. Tracings of autoradiographic densities of bound (—) and unbound (···) lanes of ethylation data of Fig. 2. The x coordinate is named by the nucleotide whose 5′ phosphate is ethylated. Vertical arrowheads, promoter positions that mark the boundaries of the ethylation responsive regions (includes phosphates at which ethylation had at least a 2-fold effect); horizontal arrows, conserved CAAT-box sequences.

LTR site, which displays no obvious palindrome, appears to bind C/EBP with the same two-fold symmetry observed in the mTf site analysis. This symmetrical pattern of contacts of C/EBP on the LTR site demonstrates that palindromic sequences are not necessary for rotationally symmetrical binding of a protein dimer.

Major Groove Contacts. The most prominent purine and phosphate contacts mapped by the alkylation interference experiments are presented on a B-DNA helix in Fig. 5*A*. Methylation at N-7 of guanine projects a methyl group into the major groove of the DNA helix; therefore, methylation interference data predict close contacts between protein residues and the major groove of the binding site. The compilation of the methylation data from both promoters suggests three major groove contact zones (Figs. 4 and 5). One zone is at the center of the dyad axis, positions 2L through 2R. Two additional contact zones are positions 6R and 6L, each of which lie a half turn of the helix away from the dyad axis in divergent directions. The limits of major groove contact can be predicted by the lack of interference at positions 7R, 8R, 7L, and 8L. Two features of the ethylation data suggest that C/EBP is also close to the major groove between the proposed contact zones. Ethylation interference maps to sites on both helix strands that rim the major groove connecting the three zones (see Fig. 4). Another distinctive feature of the ethylation data comes from comparing the boundaries of the ethylation responsive regions of the top and bottom strands (see Fig. 4). On the left side of the site, interfering ethylations on the top strand map farther from the center than do interfering ethylations on the bottom strand. This 5′ overhang pattern is also observed on the right half of the site. Due to the helical pitch and the anti-parallel configuration of B-DNA (see Fig. 5*A*), this pattern suggests major groove docking on each half-site. In conclusion, the effects of alkylation on C/EBP binding indicate continuous contacts along the major groove over a full helical turn.

DISCUSSION

We have used alkylation interference experiments to investigate the binding of C/EBP to the CAAT-box type DNA-binding site. Our primary goal was to investigate the essential DNA contacts made by the C/EBP dimer. The data demonstrate that the essential contacts display two-fold symmetry. Furthermore, contacts are in the major groove over 12–14 contiguous nucleotide pairs. Fig. 5 presents a model of the C/EBP-DNA interface. In addition to the potential purine and phosphate contacts, an undefined rotationally symmetrical protein structure is displayed that makes major groove contacts for a full helical turn. The two parts of the protein structure illustrate symmetrical domains of DNA contact but not necessarily dimer subunits. In our model, a C/EBP dimer interacts closely with the DNA helix over a full 360° to establish all of the described contacts.

Fig. 5*C* also illustrates two alternative orientations of the dimer interface of C/EBP (the leucine zipper) with respect to DNA helix structure. In one scenario, the amino termini of the leucine zipper helices are positioned over the minor groove at the dyad axis of the binding site (Fig. 5*C*, zipper a). In this case, the major groove contacts closest to the leucine zipper are in two successive major grooves on one face of the helix. Alternatively, the leucine zipper is positioned over the dyad axis in a major groove (Fig. 5*C*, zipper b).

A comparison of C/EBP contacts to well-characterized DNA–protein interfaces suggests that leucine zipper proteins bind DNA in a different manner (Fig. 6). The phosphate ethylations that interfere with DNA binding by λ repressor, for example, have two-fold symmetry but map farther from the dyad axis and show a longer 5′ overhang than is seen in the C/EBP data (17). The ethylations that interfere with λ repressor binding are consistent with the crystallographic picture of λ repressor and operator in which two helix–turn–helix structures bind two successive major grooves on one face of the DNA (19). The pattern of C/EBP contacts is

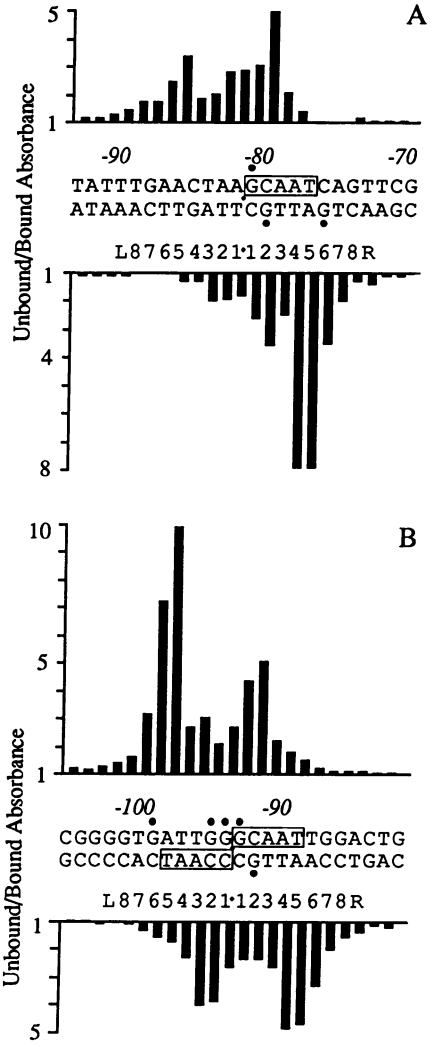

FIG. 4. Alkylation interference on LTR (*A*) and mTf (*B*) sites. For quantitative analysis, extent of ethylation interference is expressed as the relative autoradiographic density of lanes loaded with unbound and bound DNA. Histogram bars, average values of two independent experiments, are positioned 5′ to base-pair labels to indicate ethylation of the 5′ phosphate of each nucleotide. ●, positions of interfering methylations. The CAAT-box sequences are in boxes.

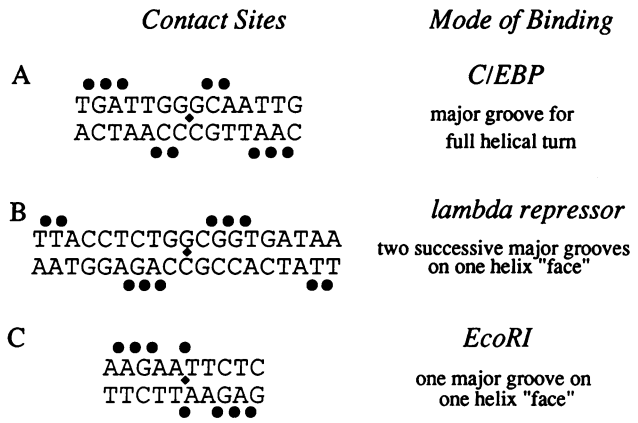

FIG. 6. Ethylation interference effects observed with C/EBP (*A*) (data from Fig. 4*B*), λ repressor (*B*) (17), and *Eco*RI (*C*) (18). ●, Strongly interfering ethylations.

also distinct from the arrangements of phosphate contacts made by the restriction endonuclease *Eco*RI. Ethylation interference analysis and the crystal structure of *Eco*RI with its binding site show that essential contacts have two-fold symmetry and map to the eight nucleotide pairs surrounding the dyad axis. These contacts are made predominantly on one face of the helix (18, 20). In contrast, the phosphate and guanine contacts made by C/EBP map farther out from the dyad axis and cannot be attained by docking on one face of the helix. As described earlier, C/EBP appears to make contacts around the DNA helix.

Vinson *et al.* (21) have recently proposed a model of C/EBP positioned on a DNA sequence that displays a palindromic GCAAT motif. The amino termini of the leucine zipper α-helices are proposed to approach the major groove at the dyad axis of the binding site. This orientation would be consistent with the model presented in Fig. 5*C* (zipper b). This model also predicts that each basic domain of a C/EBP dimer is in a hinged α-helical structure that makes DNA contacts in the major groove on one-half of a binding site. The model predicts that the C/EBP dimer makes intimate contacts on the DNA helix solely in the major groove spanning 16 bp of DNA. Our data demonstrate major groove interactions that are consistent with this model.

The symmetrical binding of C/EBP to the LTR site calls attention to the issue of the sequence-specificity of C/EBP binding. The ethylation interference data demonstrate that both halves of the LTR site mediate stabilizing contacts.

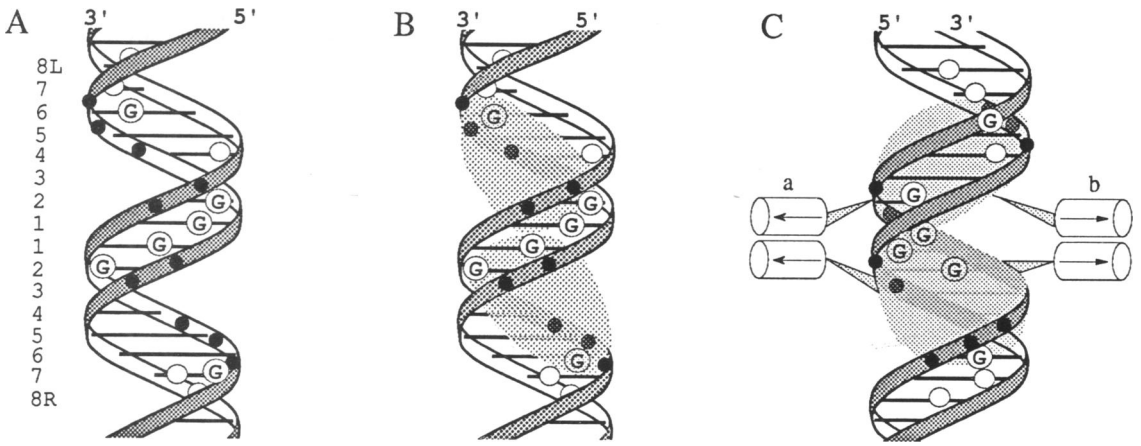

FIG. 5. Model of C/EBP–DNA interaction. ●, Phosphates at which ethylation strongly interferes with binding (greater than 3-fold effect on the mTf site); G, interfering guanine methylation (mTf and LTR sites); ○, noninterfering guanine methylation. In the view in *C*, the helix is rotated 90° with respect to the view in *A* and *B*. Lightly stippled area represents undefined protein structure. Pairs of cylinders are leucine zippers in two alternative orientations, a and b.

Furthermore, the pattern of potential phosphate contacts displays two-fold symmetry similar to that observed on the more symmetrical mTf site. Yet, the DNA sequence of the LTR site appears to have little symmetry. Within the 14 bp of the C/EBP contact zone, only two of the seven positions on the left half of the LTR site match the sequence on the right half of the site. There are two possible explanations for these results. On the one hand, the constellation of base pair functional groups could display more symmetry than we read in the simple base pair sequence. For example, N-7 on the imidazole ring of both guanine and adenine could be viewed as an equivalent H-acceptor site by a protein residue (22). Alternatively, the residues of C/EBP that are at the DNA–protein interface on the non-GCAAT half-site may make different contacts than the residues near the GCAAT half-site. This explanation would require substantial flexibility in the DNA-binding residues of C/EBP.

Whereas the data collected from the LTR site suggest possible flexibility of the DNA recognition mode of C/EBP, the tighter binding of the mTf promoter site (data not shown) stresses the potential importance of the CAAT-box sequence motif. Other relatively high-affinity binding sites for C/EBP also bear the 5′-GCAA-3′ portion of the CAAT motif of mTf on one half-site. These sites include enhancer elements of the hepatitis B virus (5′-GCAAA-3′) (11) and the mouse polyoma virus (5′-GCAAG-3′) (9). Following this reasoning, the 8-bp sequence 5′-TTGCGCAA-3′, which is a palindrome of the 5′-GCAA-3′ motif, should be an optimal site for high-affinity binding of C/EBP. Other leucine zipper/basic region proteins have been reported to also require a minimum of eight or nine nucleotide pairs for sequence-specific recognition (5, 23–25). If the central eight nucleotides of a C/EBP binding site are sufficient for sequence-specific recognition, the critical major groove and backbone contacts near positions 5R–6R and 5L–6L (Fig. 4) could be zones of sequence-independent stabilization. Stabilizing phosphate contacts also occur beyond the 6-bp recognition site of the restriction enzyme *Eco*RI (18, 20). A more quantitative analysis of C/EBP binding sites and site-directed mutagenesis of sites will be necessary to resolve the issue of the sequence-specific recognition of C/EBP.

In summary, we have shown that C/EBP binds DNA in a previously undescribed mode. The DNA determinants for the recognition of CAAT-box sites by C/EBP include 12–14 nucleotide pairs. Contacts are made in the major groove for a full turn of the helix. To accommodate all of the described contacts, C/EBP must have the flexibility to wrap around the DNA. The symmetrical binding of the C/EBP dimer to both a symmetrical and an asymmetrical binding site also suggests flexibility of C/EBP contact residues.

We thank T. Alber, G. Herrick, and B. Bass for critical comments on the manuscript. We are grateful to G. Stanley McKnight for the mTf promoter clone and to Steve McKnight and William Landschulz for the C/EBP pET expression clone. The excellent clerical assistance of Jeni Urry is also acknowledged. This work was supported by National Institutes of Health Grant GM38663.

1. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240,** 1759–1764.
2. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243,** 538–542.
3. O'Shea, E. K., Rutkowski, R., Stafford, W. F. I. & Kim, P. S. (1989) *Science* **245,** 646–648.
4. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243,** 1681–1689.
5. Turner, R. & Tjian, R. (1989) *Science* **243,** 1689–1694.
6. Agre, P., Johnson, P. F. & McKnight, S. L. (1989) *Science* **246,** 922–926.
7. Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58,** 799–839.
8. Siebenlist, U. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 122–126.
9. Johnson, P. F., Landschulz, W. H., Graves, B. J. & McKnight, S. L. (1987) *Genes Dev.* **1,** 133–146.
10. Graves, B. J., Johnson, P. F. & McKnight, S. L. (1986) *Cell* **44,** 565–576.
11. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988) *Genes Dev.* **2,** 786–800.
12. Idzerda, R. L., Behringer, R. R., Theisen, M., Huggenvik, J. I., McKnight, G. S. & Brinster, R. L. (1989) *Mol. Cell. Biol.* **9,** 5154–5162.
13. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
14. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S., Dunn, J. J. & Studier, F. W. (1987) *Gene* **56,** 125–135.
15. Fried, M. & Crothers, D. M. (1981) *Nucleic Acids Res.* **9,** 6505–6525.
16. Sun, L. & Singer, B. (1975) *Biochemistry* **14,** 1795–1802.
17. Johnson, A. (1980) Thesis (Harvard Univ., Cambridge, MA).
18. Lu, A.-L., Jack, W. E. & Modrich, P. (1981) *J. Biol. Chem.* **256,** 13200–13206.
19. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242,** 893–899.
20. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986) *Science* **234,** 1526–1541.
21. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246,** 911–916.
22. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 804–808.
23. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986) *Science* **234,** 451–457.
24. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. (1988) *Science* **242,** 1430–1433.
25. Tabata, T., Takase, H., Takayama, S., Nakayama, T. & Iwabuchi, M. (1989) *Science* **245,** 965–968.