Journal of Virology, Apr. 1993, p. 2408–2411
0022-538X/93/042408-04$02.00/0

Vol. 67, No. 4

Sequence Specificity of the Core-Binding Factor

IRENA N. MELNIKOVA,† BARBARA E. CRUTE, SHUWEN WANG,‡ AND NANCY A. SPECK*

Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755

Received 19 August 1992/Accepted 28 December 1992

The core-binding factor (CBF) binds the conserved core motif in mammalian type C retrovirus enhancers. We analyzed the phosphate contacts made by CBF on the Moloney murine leukemia virus enhancer by ethylation interference assay. The phosphate contacts span 9 bp centered around the consensus core site. To examine the sequence preferences for CBF binding, we employed the technique of selected and amplified binding sequence footprinting (T. K. Blackwell and H. Weintraub, Science 250:1104–1110, 1990). The consensus binding site for CBF defined by selected and amplified binding sequence footprinting is PyGPyG GTPy.

The core site is a sequence motif present in the enhancers of many mammalian type C retroviruses (5). It is an important genetic determinant of pathogenesis of two retroviruses, the Moloney murine leukemia virus (Mo-MuLV) and the SL3-3 MuLV. A 2-bp mutation in both of the core sites in Mo-MuLV increases the latent period of leukemia caused by the Moloney virus and shifts the specificity of the disease from thymic leukemia to erythroid leukemia (15). A 3-bp mutation in each of the four core sites in the SL3-3 MuLV renders that virus nonleukemogenic (7).

Although the core site is present in all mammalian type C retroviral enhancers, the sequences of these core sites vary between independent virus isolates. A comparison of enhancer sequences from 35 independent full-length integrated proviruses of ecotropic, xenotropic, amphotropic, and polytropic host ranges identified bases at several positions in the core site common to all enhancers (TGPyGGTN). It has been proposed that variability at the pyrimidine (TGPyG GTN) and at position N (TGPyGGTN) may account for differences in the pathogenic properties of some retroviruses (3, 4). For example, the thymotropic leukemogenic SL3-3 MuLV and the nonleukemogenic AKV MuLV contain a T and C at position N, respectively. Mutation of the SL3-3 MuLV core site (TGTGGTT) to the sequence found in the AKV MuLV core site (TGTGGTC) decreases transcription from the SL3-3 MuLV enhancer specifically in T cells (3) and alters the pathogenic properties of the SL3-3 virus (10).

Nuclear proteins that bind to distinct versions of the core site have been detected (3), as have proteins that bind to multiple core sites (3, 17–19). We have purified proteins from calf thymus nuclei that bind to the core site in the Mo-MuLV enhancer (TGTGGTA) (19). These proteins, which we call core-binding factor (CBF), appear to be analogous to the SL3-3 enhancer factor 1 (SEF1) characterized by Thornell et al. (17, 18) and the SL3-3 and AKV MuLV CBF characterized by Boral et al. (3), which binds to multiple core sites from mammalian type C retrovirus enhancers. CBF is probably also identical to the PEBP2 polyomavirus enhancer-binding proteins, which bind a core sequence (TGPyGGTPy) in the polyomavirus enhancer (8).

Here we determine the sequence specificity of DNA binding by CBF and analyze the phosphate contacts made by CBF on the DNA backbone. To determine the sequence preferences at each position of the binding site for CBF, we performed selected and amplified binding sequence (SAAB) footprinting (2). The DNA template for SAAB footprinting contained three guanines conserved in all core sites found in mammalian type C retroviral enhancers (Fig. 1) (5). These guanines were identified by methylation interference analyses as critical contacts for CBF and SEF1 binding (17, 19). The conserved guanines were flanked by randomly specified positions (N), which were in turn flanked by conserved sequences for the purpose of annealing oligonucleotide primers (primers A and B) for polymerase chain reaction (PCR); these conserved flanking sequences also contained restriction sites for *Bam*HI and *Eco*RI endonucleases. A control template that contained the corresponding sequence from the Mo-MuLV enhancer between the conserved PCR primer binding sites was synthesized. The sequences of the random oligonucleotide and primers are shown in Fig. 1. All oligonucleotides were synthesized by Operon Technology, Inc. The random and control DNA templates were purified by high-pressure liquid chromatography. A total of 100 pmol of primer A was end-labeled with [γ-^{32}P]ATP (7,000 Ci/mmol; ICN) and T4 polynucleotide kinase (New England Biolabs). Double-stranded template was generated by PCR with 100 pmol of ^{32}P-end-labeled primer A, 100 pmol of primer B, and 10 pmol of the random or control template according to the instructions provided by the manufacturer (Perkin-Elmer Cetus). Reactions were incubated at 94°C (1 min), 37°C (1 min), and 72°C (1 min) for 30 cycles and at 72°C (7 min). Double-stranded probes were purified by electrophoresis through a 12% native polyacrylamide gel and eluted from the gel by electrophoresis onto an NA45 membrane (Schleicher and Schuell) (1). The specific activity of the probe was approximately 10^6 cpm/pmol.

Electrophoretic mobility shift assays were performed as previously described, with 5 μl of affinity purified CBF and random or control probe in the binding reaction mixtures (10,000 cpm per reaction mixture) (19). The position of the protein-DNA complex in the random probe, which cannot be seen, was estimated from the position of the protein-DNA complex in the control probe (data not shown). DNA from the protein-DNA complex in the random probe was eluted from dried gels in elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% sodium

* Corresponding author.

† Present address: Institute of Biotechnology, University of Texas Health Science Center, San Antonio, TX 78245.

‡ Present address: Department of Physiology, University of California School of Medicine, San Francisco, CA 94143.

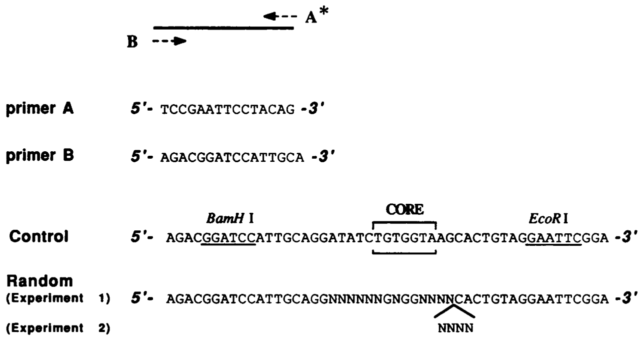
A*
B
primer A 5'- TCCGAATTCCTACAG -3'
primer B 5'- AGACGGATCCATTGCA -3'
BamH I
CORE
EcoR I
Control 5'- AGACGGATCCATTGCAGGATATCTGTGGTAAGCACTGTAGGAATTCGGA -3'
Random
(Experiment 1) 5'- AGACGGATCCATTGCAGGNNNNNNGNGGNNNNCACTGTAGGAATTCGGA -3'
(Experiment 2) NNNN

FIG. 1. Templates and primers used in SAAB analysis. The sequence of the control probe is derived from the Mo-MuLV enhancer, flanked by sequences for annealing PCR primers A and B. The asterisk indicates the location of the ^{32}P label. The locations of restriction sites in the flanking sequences are underlined. The position of the core site is indicated by the box. The sequence of the random probe used in experiment 1 (Table 1) is shown, and the locations of the four additional random positions in the probe used in experiment 2 (Table 1) are indicated below the sequence.

dodecyl sulfate [SDS]), precipitated with ethanol, and amplified again by PCR as described above, with ^{32}P-end-labeled primer A and unlabeled primer B. Following three rounds of selection and amplification, the fragments were digested with *Eco*RI and *Bam*HI, gel purified, and subcloned into the Bluescript SK$^+$ phagemid. Inserts from 61 individual subclones were sequenced (12). The results of this analysis are summarized in Table 1 (data for experiment 1).

The distribution of bases at positions 1 to 5 in the sites defined by SAAB footprinting was random in experiment 1, suggesting that positions 1 to 5 are outside the binding site for CBF. At positions 6 and 8, a pyrimidine is preferentially selected. There is strong selection for thymines at positions 11 and 12. A random distribution of bases appears at position 13, while a cytosine is preferentially selected at position 14.

Affinity-purified preparations of CBF contain multiple polypeptides that bind the core site (19). To determine whether different sequences would be selected by individual polypeptides, we repeated SAAB footprinting with several CBF polypeptides that were first fractionated by SDS-polyacrylamide gel electrophoresis, eluted from individual gel slices, and subjected to a denaturation-renaturation regimen (6, 19). The data from three independent peptides were consistent with those obtained with the mixture of affinity-purified proteins, although the percentage of thymines versus cytosines selected at positions 6 and 8 was variable (data not shown). We conclude that at least several of the polypeptides in the CBF preparation have similar binding preferences.

An extensive mutagenic analysis of the core site by Thornell et al. did not identify position 14 as a determinant for SEF1 binding (18). Thus, despite the preference for cytosine at position 14 detected by SAAB footprinting, we favor the interpretation that position 14 is outside the binding site for CBF. However, we could not exclude the possibility that position 14 is within the CBF binding site on the basis of our SAAB footprinting data. Furthermore, since position 14 was the last random position on the probe, it was possible that there are additional bases flanking position 14 which are significant for CBF binding but were not tested. To address these possibilities, we repeated the SAAB footprinting with a probe that contained four additional random positions (positions 15 to 18 [Table 1; data for experiment 2]). In this experiment the distribution of bases at position 14 is random. However, with this longer probe, preferences for a cytosine

TABLE 1. CBF binding sites selected in vitro by SAAB footprinting

Position	Nucleotide			% of sites[a]								Consensus
				Expt 1				Expt 2				
	Random[b]	Moloney[c]	Plurality[d]	G	A	T	C	G	A	T	C	
1	N	A	A	48	36	5	11	14	76	6	4	N
2	N	T	T	55	15	21	9	94	2	2	2	N
3	N	A	A	50	34	6	10	95	2	2	1	N
4	N	T	T	17	15	23	45	58	10	12	10	N
5	N	C	C	26	12	31	31	8	70	11	11	N
6	N	T	T	3	6	65	26	4	2	83	10	Py
7	G	G	G	100				100				G
8	N	T	T[e]	0	2	34	64	5	0	8	87	Py
9	G	G	G	100				100				G
10	G	G	G	100				100				G
11	N	T	T	0	2	95	3	1	2	93	2	T
12	N	A	C	0	0	92	8	2	2	69	27	Py
13	N	A	A	42	15	37	6	6	12	80	2	N
14	N	G	A	18	0	8	74	29	14	31	25	N
15	N	C	G					11	4	70	17	N
16	N	A	C					42	6	8	43	N
17	N	G	A					6	0	4	90	N
18	N	T	G					2	1	8	88	N

[a] Data for experiment 1 are percentages of sites in which the designated base appeared at each position in the binding site, from analysis of 61 independent sequences. Data for experiment 2 are percentages of sites in which the designated base appeared at each position in the binding site, from analysis of 83 independent sequences. Data obtained with affinity-purified CBF and an individual subunit of CBF renatured from an SDS-polyacrylamide gel are combined.

[b] Random oligonucleotide pool from which CBF binding sites were selected in vitro.

[c] Corresponding sequence from the Mo-MuLV enhancer.

[d] The nucleotide most commonly found at each position in mammalian type C retroviral enhancers (5).

[e] Thymine is found at position 8 in 89% of mammalian type C retroviral enhancers, and cytosine is found in the remaining 11%.

at positions 17 and 18 and a guanine at positions 2 and 3 are apparent. We cannot explain these preferences, nor can we explain the variability with which they occur in different probes. We speculate that perhaps the bases that are preferentially selected at these positions were overrepresented in the original oligonucleotide pool.

When we analyze the data from both probes and include only those bases that are consistently selected in defining a consensus binding site for CBF, we derive the sequence PyGPyGGTPy. This sequence is within the consensus high-affinity binding site for SEF1 (TTTGCGGTTA) defined by Thornell et al. (18).

The phosphate contacts for CBF were identified by ethylation interference analysis (13). Phosphate triester formation with *N*-ethyl-*N*-nitrosourea adds an ethyl group and removes a negative charge from the DNA backbone. Inhibition of CBF binding by phosphate ethylation can be due to steric interference or disruption of an electrostatic interaction between CBF and the DNA backbone. Probes were prepared from a plasmid which contains *Hae*III$_{7949}$-*Hae*III$_{7994}$ sequences from the wild-type Mo-MuLV enhancer (numbered as described by Weiss et al. [20]) subcloned into the *Sma*I site of SP64 (14). The *Hae*III$_{7949}$-*Hae*III$_{7994}$ probe was prepared by cutting the SP64 polylinker at either the *Eco*RI or the *Bam*HI site, dephosphorylating the 5′ end with calf intestinal phosphatase (Boehringer Mannheim Biochemicals), end labeling with [γ-^{32}P]ATP, and recutting with either *Bam*HI or *Eco*RI. The probes were modified on phosphates with *N*-ethyl-*N*-nitrosourea (Sigma), as described by Siebenlist and Gilbert (13). Binding reaction mixtures (75-μl total volume) contained 100,000 cpm of end-labeled ethylated probe and 10 μl of affinity-purified CBF. The binding and electrophoresis conditions were identical to those described previously for methylation interference (19). Purified DNA from the protein-DNA complex and free DNA bands from the native polyacrylamide gel were cleaved by alkali (13). The cleavage products were analyzed by electrophoresis through a 15% polyacrylamide–7 M urea sequencing gel (Fig. 2A).

The methylation and ethylation interference data are summarized in Fig. 2B. CBF contacts six phosphates on the plus strand and four phosphates on the minus strand of the binding site. These phosphates map primarily to one face of the DNA helix (we designate this the front face). The phosphate contacts at positions 8 to 10 on the plus strand wind around towards the back face of the DNA helix. Positions 6 to 12, at which sequence preferences were detected by SAAB footprinting, are within the area of phosphate contacts.

The major groove between the two areas of phosphate contacts on the front face of the helix contains the three guanines contacted by CBF and SEF1 (positions 7, 9, and 10) (18, 19). Base contacts in the minor groove were also detected by methylation interference at positions 11 to 13 (19). We propose that CBF contacts bases primarily in the major groove between the two areas of phosphate contacts on the front face of the helix. The contacts to adenines in the minor groove could be generated by an arm of the polypeptide reaching into the minor groove, as seen for the engrailed and MATα2 homeodomain proteins (9, 21). Alternatively, methylation at the N-3 atom of adenines 11 to 13 in the minor groove could induce a conformational change in the DNA which prevents CBF binding, or it could interfere with a necessary conformational change induced in the DNA as a result of CBF binding, such as narrowing of the minor groove.

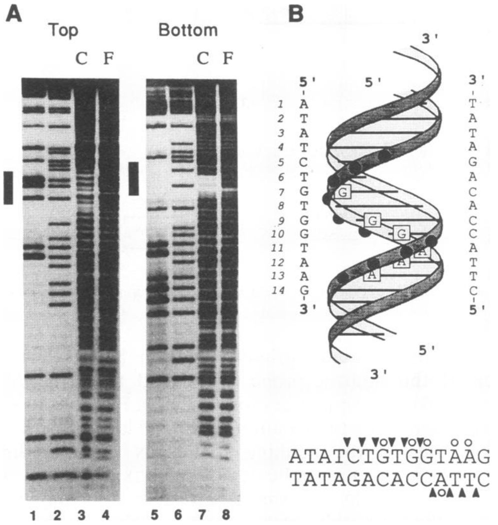

FIG. 2. Ethylation interference of CBF on the Mo-MuLV enhancer. (A) Ethylation interference of CBF on a 45-bp region from the Mo-MuLV enhancer. Lanes 1 to 4, top (plus) strand of the Mo-MuLV enhancer; lanes 5 to 8, bottom (minus) strand. Lanes 1 and 5 and lanes 2 and 6 represent G and G+A chemical sequencing tracts, respectively. Lanes 3 and 7 correspond to the protein-DNA complex, and lanes 4 and 8 correspond to the free DNA bands from the mobility shift gel. Vertical boxes indicate the location of the core site. (B) Summary of ethylation and methylation interference results. In the upper portion is the B-form DNA helix showing the phosphate and purine base contacts by CBF. Open squares with letters (G or A) in the ladder represent bases in the CBF binding site identified by methylation interference (19). Solid circles on the DNA backbone represent phosphates identified as contacts for CBF by ethylation interference. The numbers correspond to the positions designated in Table 1. In the bottom portion, the phosphate contacts are shown by arrowheads and the purine contacts are shown by open circles above and below the sequence.

How does the consensus core site defined by SAAB footprinting compare with core sites found in viral and cellular enhancers? In 34 of 35 mammalian type C retroviral enhancers, the bases found at positions 1 to 5 are 5′-ATATC-3′ (5). Since these sequences appear to be outside the binding site for CBF defined by SAAB footprinting, selective pressures other than CBF binding must maintain the conservation of bases at positions 1 to 5 in retroviral enhancers. At position 6, a thymine is present in 100% of type C retroviral enhancers and in 76% of the sites selected by CBF in vitro. Core sites in retroviral enhancers contain a pyrimidine at position 8, as do the core sites identified by SAAB. However, cytosine occurs at position 8 in 77% of the core sites selected by SAAB in vitro but in only 11% of retroviral enhancers. A thymine at position 11 is found in 35 of 35 retroviral enhancers (5), as well as in the consensus site defined by SAAB footprinting. At position 12, a thymidine appears in 5 of 35 retroviral enhancers (14%) but much more frequently in the sites selected by SAAB footprinting (79%). In fact, an adenine is more often found at position 12 in mammalian type C retroviral enhancers (31%), including the Mo-MuLV enhancer, yet an adenine at position 12 was rarely selected in vitro. The frequency at which a thymine appears at position 12 in the in vitro selected sites suggests that CBF binds with a higher affinity to core sites with a thymine at this position. This is consistent with data from Thornell et al. for the protein SEF1, which we believe may be identical to CBF on the basis of its tissue distribution and sequence specificity (17–19). We note that several of the highly thymotropic MuLVs including SL3-3 and Gross passage A, have a thymine at position 12. What advantage is conferred to the Moloney virus by an adenine at position 12 is unclear. We note, however, that the SAAB analysis may detect extremely small differences in binding specificity and that these have uncertain, if any, functional significance.

Core sites with the sequence TGTGGTT are also found in functionally important regions of the enhancers from several cellular genes transcribed in T cells, including the T-cell receptor β-, γ-, and δ-chain genes, the immunoglobulin μ-chain gene, and the CD3 ε-chain gene (11, 16, 18, 19). A mutation in the core site in the T-cell receptor δ-chain enhancer significantly attenuates transcription of this gene in T cells (11).

We thank Barbara Graves for technical advice, Laura Reisman for technical assistance, and Ranjan Sen and Paul Gardner for critically reading the manuscript.

This work was supported by Public Health Service grant CA51065-01A1 from the National Cancer Institute, by a Cancer Center CORE grant (CA23018) from the National Cancer Institute to the Norris Cotton Cancer Center, and by the Leukemia Research Foundation. N.S. is a recipient of a Junior Faculty Research Award from the American Cancer Society, and S.W. is an awardee of a predoctoral fellowship from the Norris Cotton Cancer Center at the Mary Hitchcock Hospital.

REFERENCES

1. **Baldwin, A.** 1988. Methylation interference assay for analysis of DNA-protein interactions, p. 12.3.2–12.3.3. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
2. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of myoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
3. **Boral, A. L., S. A. Okenquist, and J. Lenz.** 1989. Identification of the SL3-3 virus enhancer core as a T-lymphoma cell-specific element. J. Virol. **63:**76–84.
4. **Clark, S. P., and T. W. Mak.** 1982. Nucleotide sequences of the murine retrovirus Friend SFFVp long terminal repeats: identification of a structure with extensive dyad symmetry 5′ to the TATA box. Nucleic Acids Res. **10:**3315–3330.
5. **Golemis, E. A., N. A. Speck, and N. Hopkins.** 1990. Alignment of U3 region sequences of mammalian type C viruses: identification of highly conserved motifs and implications for enhancer design. J. Virol. **64:**534–542.
6. **Hager, D. A., and R. R. Burgess.** 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of Escherichia coli RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. Anal. Biochem. **109:**76–86.
7. **Hallberg, B., J. Schmidt, A. Luz, F. S. Pedersen, and T. Grundström.** 1991. SL3-3 enhancer factor 1 transcriptional activators are required for tumor formation by SL3-3 murine leukemia virus. J. Virol. **65:**4177–4181.
8. **Kamachi, Y., E. Ogawa, M. Asano, S. Ishida, Y. Murakami, M. Satake, Y. Ito, and K. Shigesada.** 1990. Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. J. Virol. **64:**4808–4819.
9. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
10. **Lenz, J. (Albert Einstein College of Medicine).** 1992. Personal communication.
11. **Redondo, J. M., J. L. Pfohl, and M. S. Krangel.** 1991. Identification of an essential site for transcriptional activation within the human T-cell receptor δ enhancer. Mol. Cell. Biol. **11:**5671–5680.
12. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
13. **Siebenlist, U., and W. Gilbert.** 1980. Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77:**122–126.
14. **Speck, N. A., and D. Baltimore.** 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. Mol. Cell. Biol. **7:**1101–1110.
15. **Speck, N. A., B. Renjifo, E. Golemis, T. N. Fredrickson, J. W. Hartley, and N. Hopkins.** 1990. Mutation of the core or adjacent LVb elements of the Moloney murine leukemia virus enhancer alters disease specificity. Genes Dev. **4:**233–242.
16. **Spencer, D. M., Y.-H. Hsiang, J. P. Goldman, and D. H. Raulet.** 1991. Identification of a T-cell-specific transcriptional enhancer located 3′ of Cγ1 in the murine T-cell receptor γ locus. Proc. Natl. Acad. Sci. USA **88:**800–804.
17. **Thornell, A., B. Hallberg, and T. Grundström.** 1988. Differential protein binding in lymphocytes to a sequence in the enhancer of the mouse retrovirus SL3-3. Mol. Cell. Biol. **8:**1625–1637.
18. **Thornell, A., B. Hallberg, and T. Grundström.** 1991. Binding of SL3-3 enhancer factor 1 transcriptional activators to viral and chromosomal enhancer sequences. J. Virol. **65:**42–50.
19. **Wang, S., and N. A. Speck.** 1992. Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. Mol. Cell. Biol. **12:**89–102.
20. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1985. RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo.** 1991. Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67:**517–528.