A Novel Matrix Attachment Region DNA Binding Motif Identified Using a Random Phage Peptide Library*

Bingcheng Wang, Liliane A. Dickinson, Erkki Koivunen, Erkki Ruoslahti, and Terumi Kohwi-Shigematsu†

From the Cancer Research Center, La Jolla Cancer Research Foundation, La Jolla, California 92037

SATB1 is a novel DNA-binding protein which is predominantly expressed in thymocytes. This protein binds to the minor groove specifically recognizing an unusual DNA context exhibited by a specific MAR region with strong base-unpairing propensity. A phage library displaying nonamer random peptides without any built-in structure was used to identify a MAR binding motif of SATB1. One predominant cyclic peptide C1 of CRQNWGLEGC selected by a MAR-affinity column showed 50% identity with a segment in SATB1 (amino acids 355–363). Replacement of the C1 similarity segment in SATB1 by a random amino acid sequence or its truncation resulted in more than 80% reduction in MAR binding. In contrast, replacement of the same SATB1 segment with the C1 peptide restored full MAR binding activity and specificity as the wild-type protein. Single amino acid mutation of the conserved Arg or Glu residue to Ala greatly reduced MAR binding activity. Taken together our data show that a nine-amino acid sequence in SATB1 represents a key MAR binding motif. Phage display may provide a general tool for rapid identification of DNA binding peptide motifs.

Eukaryotic chromatin is organized into loop domains, which may have both structural and functional roles such as differential gene expression and replication (1–3). It is believed that the chromatin loops are anchored to the nuclear matrix at specific DNA regions that exhibit high affinity to the nuclear matrix. These matrix-associated regions or MARs (2, 3) are often found near enhancer-like regulatory sequences (9–12). Common at the boundaries of transcription units (4–8) and often found near enhancer-like regulatory sequences (9–12). Recent evidence shows that MARs play a role in tissue-specific gene expression. MARs associated with the immunoglobulin μ heavy chain locus have been shown to be essential for transcription of a rearranged μ gene in transgenic B lymphocytes (13). Also, MARs are binding targets for a cell type-specific MAR-binding protein, SATB1. SATB1 is primarily expressed in thymus (14), and because it is expressed during specific stages of T-cell development, it is believed to be involved in thymocyte maturation and differentiation.

SATB1 is a novel type of DNA-binding protein that recognizes a specific sequence context in which one strand exclusively consists of mixed A, T, and C nucleotides (ATC sequences). Clustered ATC sequences commonly found in MARs have a strong tendency to unwind by extensive base unpairing (15). The unwinding property of MARs confers high affinity binding to the nuclear matrix and is essential for transcription enhancing activity of such sequences (16). SATB1 does not bind to sequences that are similarly AT-rich but lack the unwinding capability. SATB1 binding is highly specific, but exhibits an unusual mode of DNA recognition; it interacts with the minor grooves of its target sequences while making little contact with the bases (14). A 150-amino acid segment of SATB1 was recently identified as the MAR binding domain (17). The concomitant presence of both NH2- and COOH-terminal arms of this segment is necessary for full binding activity, but the DNA contact sites remain to be determined.

We have taken an unconventional approach in delineating the DNA binding sites within SATB1 using a random peptide bacteriophage display library. Bacteriophage display is a powerful tool to study protein-target molecule interactions (18, 19). In this system, hundreds of millions of random peptides are expressed on the surface of bacteriophage as fusion protein libraries, and ligands for various purposes can be selected from them. Peptide ligands for proteins (20–24), antibodies (25), and enzymes (26) have been identified in this manner. We speculated that phage display libraries could also represent a vast source of DNA binding motifs. We report here that a random peptide bacteriophage library without any built-in structures can be used to affinity-select specific MAR binding peptides. Based on the sequence similarities between the selected peptide and a native MAR-binding protein SATB1, we were able to define a nine-amino acid segment in SATB1 as a key MAR binding motif for this protein. Deletion of this segment markedly reduced the MAR binding activity of SATB1. In the context of SATB1, this motif together with other components in SATB1 confers unique binding specificity to the AT-rich sequences with strong unwinding potential.

EXPERIMENTAL PROCEDURES

MAR Probes and Affinity Column—Multimeric wild-type MAR DNA (25)9 (top strand of the monomer sequence: 5′-CTTTAATTTTCT-ATATATTAGAAC-3′) and mutant MAR DNA (24)9 (top strand of the monomer sequence: 5′-CTTTAATTTTCTAGTTAGAAC-3′) were generated as described (14). The differences between the wild-type and mutated MARs are underlined. To prepare affinity column, concatemers of the wild-type MAR were coupled to CNBr-activated Sepharose 6MB (Pharmacia Biotech Inc.) at 200 μg/ml bed volume as described (27). A control column was made by coupling sonicated herring sperm DNA (average molecular weight about 200 base pairs) to Sepharose 6MB at the same concentrations.

Construction of CXb Library—The details for constructing CXb library (C, cysteine; X, any amino acids) have been described elsewhere (28). Briefly, an oligonucleotide corresponding to the sense strand of CXb insert was synthesized and amplified with primers complementary to the surrounding constant regions. The purified polymerase chain reaction product was cloned into the amino terminus coding region of the C1 peptide and an affinity-selected MAR-binding protein SATB1, we were able to define a nine-amino acid segment in SATB1 as a key MAR binding motif for this protein. Deletion of this segment markedly reduced the MAR binding activity of SATB1. In the context of SATB1, this motif together with other components in SATB1 confers unique binding specificity to the AT-rich sequences with strong unwinding potential.

1 The abbreviations used are: MAR, matrix attachment DNA; TU, transducing units.
2 L. A. Dickinson, I. deBelle, and T. Kohwi-Shigematsu, unpublished results.

* This work was supported by NIH Grants CA 28896 (to E. R.) and CA 39681 (to T. K.-S.), Tobacco-related Disease Grant 4IT0147 (to E. R.), and Cancer Center Support Grant Grant CA 30199. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† To whom correspondence should be addressed. Tel.: 619-455-6480 (ext. 681); Fax: 619-453-6217; E-mail: terumiks@ljcrf.edu.
‡ To whom correspondence should be addressed. Tel.: 619-453-6217; E-mail: terumiks@ljcrf.edu.
obtained and cultured in 2 liter of medium. Phage seeded into media were precipitated, resuspended, and stored at 4 °C as a primary library.

Phage Selection on a MAR Affinity Column—Our selection strategy combined procedures for biopanning (18, 19) and methods for purifying DNA-binding proteins (29, 30). In the first round of panning, about 10^{11} transducing units (TU) of the CX_{9} library were incubated with 0.2 ml of the concatemerized MAR Sepharose (5' end ODN 46) for 3 h at room temperature in DNA binding buffer (DBB: 25 mM HEPES-KOH, pH 7.6, 12.5 mM MgCl_{2}, 0.1 mM KCl, 20% glycerol, and 0.1% Nonidet P-40) plus 0.1% bovine serum albumin and 10 μg/ml poly(dI-dC) in a final volume of 0.5 ml. The unbound phage were removed, and the beads were washed three times with 1 ml of DBB in a microcentrifuge tube. After transfer to a 1.5-ml microcolumn (Bio-Rad), the beads were subjected to three 1-ml washes. The bound phage were then eluted successively with 0.2 ml of DBB containing increasing concentrations of KCl: once each with buffers containing 0.1 M, 0.4 M, and 0.7 M KCl and three times each with buffers containing 1.0 M and 2.0 M KCl. The eluted phage fractions were titrated, pooled, and amplified. The second to the fifth rounds of selection were carried out under the same conditions except that the input phage and incubation time were gradually reduced to 10^6 TU and 20 min, respectively. Only phage amplified from the Z1.0 and Z2.0 eluates were used in the sixth panning where the KCl concentration in the binding buffer was raised to 0.4 M. In the last selection, the KCl concentration was further raised to 0.7 M. DNA from individual phage clones was extracted and sequenced as described (18, 19).

Fig. 1. Shift in the phage elution profile following multiple rounds of selection on a MAR affinity column. Results from the first (●), fourth (■), and seventh (○) round of selection are shown. Phage from a nonamer random peptide display library were bound to a MAR DNA affinity column as described under “Experimental Procedures” and eluted with buffers containing the indicated final concentrations of KCl. The yield was estimated from the number of tetradecylamine-resistant colonies and was following infection of K91 Kan strain of E. coli cells; the y axis shows the yield from each fraction expressed as the percent of total TU recovered.

Table I lists amino acid sequences displayed on these phage and their MAR binding activity determined by a filter binding assay. After seven rounds of selection, a single peptide CRQNKGLLEG (C1) predominated over 60% (26/42) of all the phage clones sequenced and exhibited strong MAR binding activity. Sequence alignment revealed a similarity of C1 with SATB1 at the NH_{2}-terminal (four identities and one similarity among 8 residues) and COOH-terminal arms (two identities and one similarity among 8 residues) of the MAR binding domain (Fig. 2 and Ref. 17). Two other clones (C2 and C3) were similar to C1 and SATB1 in that they both contained charged residues at the conserved positions. Clone 4 (C4) was identified four times and showed some similarities to the NH_{2}-terminal arm of SATB1 adjacent to the C1 similarity region, but not to the COOH-terminal arm (not shown). The remaining clones, including clone 5 (C5) which was repeated three times, did not show discernible similarities to SATB1. Except for C1 and C4, all other clones showed low levels of binding but still above those of control phage.

All the clones had a second cysteine, suggesting a preference for cyclized peptides. The conformational constraint provided by the disulfide bond appeared to be important for the binding, as reduction of the bond greatly decreased MAR binding by C1 (Fig. 3A) and by other clones (not shown) in the filter binding assay. To test the specificity of MAR binding by the predominant C1 clone, an unrelated DNA sequence was used as a probe, and no binding was detected (Fig. 3B). A 100-fold molar excess of unlabeled wild-type MAR completely inhibited C1 binding, whereas nonspecific herring sperm DNA had no effect at much higher concentrations (Fig. 3C). However, C1 binding was also inhibited by unlabeled mutated MAR (5'-TCTTTAATTTCATCTTGTTAGGAA-3') which has lost the base pairing capability and is not bound by SATB1, even though it is still AT-rich (14). As described below, it appears that the context of the whole DNA binding domain is required for the C1 similarity region in SATB1 to distinguish between...
isotopic MAR probe (Fig. 4). In contrast, a deletion mutant lacking amino-terminal 21 residues (MD-Δ367) and a protein produced by inserting a random peptide (MD-XX) had greatly reduced MAR binding activities (Figs. 4 and 5). Furthermore, like MD-wt, MD-C1 chimera discriminated between the wild-type and mutated probes, since a 200-fold molar excess of unlabeled wild-type MAR completely inhibited binding, whereas the mutated MAR had virtually no effect (Fig. 5). Thus, MD-C1 confers specific binding to the AT-rich sequence with high unwinding capability.
The amino acids in the C1 peptide critical for MAR binding were studied by point mutations. Simultaneous mutations of the three conserved residues Arg, Asn, and Glu to Ala in MD-C1 (mut RNE-AAA) reduced the affinity of the resultant chimera for MAR to the level of the deletion mutant MD-Δ367 (Fig. 4, A and B). Mutation of either Arg (mut R-A) or Glu (mut E-A) alone was sufficient to reduce binding to the same level, whereas a mutation in Asn (mut N-A) had an intermediate effect. These results show that specific amino acids, Arg and Glu, that are conserved among C1, the NH2-terminal arm and COOH-terminal arm of the SATB1 MAR binding domain are critical for MAR recognition. Thus, consistent with the predominance of C1 phase after stringent selection, results from peptide swapping, point mutations, and deletion mutations all suggest that the C1 similarity region in SATB1 is a key DNA binding site.

**DISCUSSION**

Using a random peptide phage library without any built-in structure, we have isolated a predominant peptide (C1) of eight amino acids which is similar to the sequences found in SATB1 located within the NH2- and COOH-terminal regions of the 150-amino acid MAR binding domain. Several lines of evidence suggest that this peptide represents a MAR binding motif. First, phage displaying this peptide can directly bind both to the MAR immobilized on an affinity column and to the radio-labeled MAR in solution. More importantly, it could functionally replace the similar native sequence in a chimeric SATB1 molecule (32–34).

Our results demonstrate that small peptides displayed on the surface of filamentous phage can have high affinity for specific DNA sequences and can allow identification of a specific DNA binding motif within such proteins. Recently, three zinc fingers of the Zif268 protein were expressed on the phage surface, and randomization of the DNA contacting residues was used to select for zinc fingers with new DNA binding specificities (36–38). Rather than using long peptides with built-in structures, we have demonstrated that random short peptide phage library can be effective in searching for potential DNA binding sites in DNA-binding proteins for which the DNA binding domains are unknown or have not been completely characterized. Structure/function studies of DNA binding peptides identified from peptide libraries may present new opportunities for drug discovery.

**REFERENCES**

1. Gross, D. S., and Garrard, W. T. (1987) Trends Biochem. Sci. 12, 293–297
2. Gasser, S. M., and Laemmli, U. K. (1987) Trends Genet. 3, 16–22
3. Phi-Van, L., and Straling, W. H. (1990) Trends Biochem. Sci. 15, 363–365
4. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 6, 154–158
5. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 32, 363–365
6. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 6, 154–158
7. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 32, 363–365
8. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 6, 154–158
9. Phi-Van, L., and Straling, W. H. (1990) Trends Genet. 32, 363–365
10. Gasser, S. M., and Laemmli, U. K. (1986) Trends Biochem. Sci. 12, 293–297
11. Cockerill, P. N., Yuen, M.-H., and Garrard, W. T. (1987) Trends Biochem. Sci. 12, 293–297
12. Jarnum, A. P., and Higgs, D. R. (1988) EMBO J. 7, 2337–2344
13. Forrester, W. C., van Genderen, C., J. enwein, T., and Grosschedl, R. (1994) Science 263, 1221–1225
14. Dickinson, L. A., joh, T., Kohwi, Y., and Kohwi-Shigematsu, T. (1992) Cell 70, 631–645
15. Kohwi-Shigematsu, T., and Kohwi, Y. (1990) Biochemistry 29, 3551–3560
16. Bode, J., Kohwi, Y., Dickinson, L., joh, T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. (1992) Science 255, 195–197
17. Nakagomi, K., Kohwi, Y., Dickinson, L. A., and Kohwi-Shigematsu, T. (1994) Mol. Cell. Biol. 14, 1852–1860
18. Scott, J. K., and Smith, G. P. (1990) Science 249, 386–390
19. Smith, G. P., and Scott, J. (1992) Methods Enzymol. 217, 228–257
20. Kovalunen, E., Guy, D. A., and Rusdailah, E. (1993) J. Biol. Chem. 268, 20205–20210
21. Blond-Elguindi, S., Cwirla, S., Dower, W., Lipshutz, R., Sprang, S., Sambrook, J., and Gething, M.-J. (1993) Cell 75, 717–728
22. Kovalunen, E., Wang, B., and Rusdailah, E. (1994) Cell 74, 373–380
23. Sparks, A. B., Quilliam, L. A., Thorn, J. M., Der, C. J., and Kay, B. K. (1994) J. Biol. Chem. 269, 23853–23856
24. Chiaide, C., C. vanweinchenko, Y., South, V., Searfoss, G. H., French, S., Wolf, A., Ricca, G., and Jaye, M. (1994) J. Biol. Chem. 269, 20434–20439
25. Kay, K., Adey, N., Hs, S., Manfredi, J., Mataragnon, A., and Fowlkes, D. (1993) Gene (Amst.) 128, 59–65
26. Smith, G. P., Schultz, D. A., and Ladbury, J. E. (1993) Gene (Amst.) 128, 37–41
27. Dickinson, L. A., and Kohwi-Shigematsu, T. (1995) Mol. Cell. Biol. 15, 465–465
28. Kovalunen, E., Wang, B., and Rusdailah, E. (1995) Biochemistry 34, 265–270
29. Kadonaga, J. T., and Tjian, R. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 5893–5893
30. Kadonaga, J. T. (1991) Methods Enzymol. 208, 10–23
31. O’Neil, K., Hoess, R., and Bode, J., Kohwi, Y., and Kohwi-Shigematsu, T. (1992) Proteins Struct. Funct. Genet. 14, 509–515
32. Stucka, R., and Feldmann, H. (1990) FEBS Lett. 263, 223–225
33. Reddy, P., and Hahn, S. (1991) Cell 65, 349–359
34. Nikolov, D., Hs, W., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.-H., Roeder, R., and Burley, S. (1992) Nature 359, 1–6
35. Faye, E., and Pennan, S. (1988) Proc. Natl Acad. Sci. U.S.A. 85, 121–125
36. Rebar, E. J., and Pabo, C. O. (1994) Science 263, 671–673
37. Jameson, A. C., Kim, S.-H., and Wells, J. A. (1994) Biochemistry 33, 5689–5695
38. Choo, Y., and Klug, A. (1994) Proc. Natl Acad. Sci. U.S.A. 91, 11163–11168

3 M. Sikorska, C. Carson, S. Pandey, P. R. Walker, and T. Kohwi-Shigematsu, manuscript in preparation.