Cloning and characterization of a single-stranded DNA binding protein that specifically recognizes deoxycytidine stretch

Keizo Ito, Kenzo Sato and Hideya Endo*
Department of Molecular Biology, School of Life Sciences, Faculty of Medicine, Tottori University, Yonago 683, Japan

Received October 1, 1993; Revised and Accepted December 6 1993

ABSTRACT

We previously identified a G-rich silencer element involved in negative regulation of catalase gene expression in some hepatoma cells (Mol. Cell. Biol., 1992, 12, 2525-2533). To study a nuclear binding protein for this element, we screened cDNA libraries from a rat ascites hepatoma cell line by binding with a synthetic oligonucleotide probe and obtained several clones. One of them, designated SW, was studied in detail. A clone (SW2) of this series contained a near full length cDNA encoding a putative peptide with 463 amino acid residues. We isolated this peptide as a fusion protein. It was found that the protein strongly bound to the C-stretch of the DNA sequence in a single strand specific fashion, but absolutely did not to G-rich sequence. The protein bound weakly to the corresponding double-stranded DNA as well as to C-rich RNA sequence. This protein, though not the expected one, was found to be a novel protein whose DNA binding domain was located on the region containing at least 75 amino acid residues of the carboxyl terminus. A proline rich region was also observed in the middle part of the protein. Northern blot profiles indicated extensive and slight expression of both 2.0 kb and 2.7 kb mRNA species in some hepatoma cell lines and in the rat liver, respectively.

INTRODUCTION

In the light of the classical observation that catalase activity is markedly reduced in various tumor cells, we have studied this phenomenon from the aspects of gene regulation using hepatoma cell lines. It was found that 1) the phenomenon was due to transcriptional repression, 2) one of the elements located in 5'-flanking region of the gene strongly suppressed the catalase promoter activity in vivo and in vitro, 3) the core sequence of this silencer element was TGAG from a gel shift competition assay, and 4) an approximately 35 kDa nuclear protein bound to the silencer element was present in hepatoma cells but undetected in rat liver cells (1).

These experimental results inevitably led us to the isolation and identification of the 35 kDa binding protein. One of the methods we adopted for this purpose was to obtain a cDNA clone encoding the protein. During the course of this experiment, we isolated several cDNA clones, one of which was found to code for an approximately 51 kDa, single-stranded DNA binding protein that specifically recognized the dC-stretch. It is therefore possible that this clone is not the expected one. Nevertheless, the reason why we were interested in it was that 1) the protein encoded turned out to be a novel one, 2) the protein was capable of binding, though weakly, to the double-stranded silencer element as well, and 3) the protein was strongly expressed in hepatoma cells across species, but slightly in the rat liver, as we expected in the screening. These properties of the protein motivated us to further study its structure and function. Recently, a number of reports concerning the sequence-specific, single-stranded DNA binding proteins in mammalian cells have accumulated. Of these proteins, those which recognize the pyrimidine-rich tracts are of interest, since there has been increasing evidence suggesting their involvement in replication (2-5), recombination (6,7), and transcription (8-14) of certain genes.

In this context, the dC-stretch specific, single-strand binding protein identified in the present study seems to be worth noticing. This paper describes cloning and characterization of this protein.

MATERIALS AND METHODS

Cell lines

Rat ascites hepatoma cell lines, AH66 and Ac2F, were supplied from the Cell Bank of the Japanese Cancer Research Resources. AH66 is a subline derived from Yoshida ascites hepatoma cell lines, and Ac2F is a spontaneously transformed cell line of the rat hepatocyte origin. Human hepatoma cell line, HepG2 is a gift from Dr M. Namba at Okayama University. All of these cell lines were maintained in Dulbecco modified minimal essential medium supplemented with 10 % fetal calf serum.

* To whom correspondence should be addressed
Cloning of cDNA encoding DNA binding protein

cDNA libraries of a rat ascites hepatoma cell line prepared in λgt11 and λZAP vectors were kindly supplied by Dr M. Yamamoto at National Defense Medical College. Two oligonucleotides, 5'-GGTTTGGGGAGGGGAGGAGG-3' (the sense strand of the rat catalase gene silencer core sequence) and 5'-AAACCCCTCCCCCC-3' (the antisense strand) were synthesized on Cyclon Plus DNA Synthesizer (Millipore Co.), were labeled with \([\gamma^{32}\text{P}]\text{ATP}\) at the 5'-end, annealed to each other and ligated to make a tandem repeated DNA segment. Using this segment as a probe, the λgt11 library was screened by the Southwestern method described by Singh et al. (15) with a minor modification. Because of the lack of the 5' region of the first SW clone isolated (SW1), the λZAP cDNA library was screened by plaque hybridization by probing with an EcoRI fragment from the clone to obtain full length cDNA clones. The probe was labeled with \([\alpha^{32}\text{P}]\text{dATP}\) by the random primer DNA labeling procedure. Nucleotide sequences of the clones were determined by the dideoxynucleotide chain termination method (16) using a T7 DNA polymerase sequencing system.

Preparation of the fusion protein

The coding region of SW clones and their deletion mutants with truncation in various length at the amino terminus (SW3, SW4, SW11, SW12, SW13, SW14, SW15 and SW16) or the carboxyl terminus (SW17, SW18, SW19 and SW20) were inserted to the polylinker site of an expression vector, pMal-cRI (New England Biolabs, Inc.), in frame. Physical maps of these described by Bowen et al. (17) and Kyesse-Andersen (18) using the antisense oligonucleotide 5'-end labeled with \([\gamma^{32}\text{P}]\text{ATP}\) as a probe.

Slot blot method. MBP-fusion proteins purified from the lysate by amylose resin were fixed on nitrocellulose membrane using a slot blot apparatus (BioRad Co). DNA binding reaction was performed in the presence and absence of cold competitors for 1 hr at room temperature. The membrane was then washed under the same conditions as the Southwestern blot analysis and subjected to autoradiography.

Northern blot analysis

RNA was isolated from rat organs and hepatoma cell lines as described by Chomczynski and Sacchi (19). The Northern blot analysis was carried out according to Goldberg (20).

RESULTS

Cloning of cDNA encoding DNA binding protein

In an attempt to isolate a cDNA clone encoding the 35 kDa nuclear binding protein for the catalase silencer element, about 3 x 10^6 plaques of the Agt11 cDNA library was screened by the Southwestern assay using a synthetic oligomer of the G-rich core sequence of the silencer as a probe. One of several clones thus obtained was confirmed by secondary and tertiary screening. This clone, designated SW1, was found to contain a fragment 690 bp in length, comprising a part of coding frame lacking ATG codon and 3'-non coding region including one poly A signal site and 27 bp poly A tail (Fig. 1). Hence, in order to obtain its full length clone, we again screened 1.5 x 10^6 plaques of the λZAP cDNA library by plaque hybridization with this fragment and isolated 14 positive clones. One of them, designated SW2, was found to have an about 2.6 kb insert with a near full length size. Fig. 1 shows the nucleotide sequence of SW2 cDNA and the predicted amino acid sequence.

The primary structure of SW2 clone

The coding region. In the nucleotide sequence of SW2 cDNA, the methionine codon denoted as position 1 was deduced to represent the translation start site for the following reasons. First, this ATG codon was located on the same open reading frame as SW1 cDNA. Second, the ATG codon was surrounded by sequence features typical of initiation methionine in eucaryotic cells. As shown in 85 % of bona fide start codons, the nucleotide at position -3 is an A, and the nucleotide at position +4 is usually a G for start codon (21). This rule was also followed well by the ATG codon at position 1, but not by those at position +79 and +124 in the above nucleotide sequence of SW2 cDNA. Third, ATG codons, in the upstream region of the position -12 are out of question since the stop codon TAG was located at position -12 on the same frame as the ATG codon at position 1. SW2 cDNA covered complete coding frame comprising 1389 nucleotides with a stop codon, TAA at position +1390.

SW2 cDNA was estimated to encode a 51 kDa peptide comprising 463 amino acid residues. In addition, the following features were noted upon analysis of SW2 clone. The predicted amino acid sequence contained 13.4 % basic residues, one potential N-glycosylation site (Asn-Ala-Ser) (22) at position +217, and one nuclear localization signal (Lys-Arg-Ser-Arg) (23) at position +100. There were three repeats of hydrophobic and basic motif (class I repeat) and two repeats of a 14 amino acid sequence (class II repeat). Between both sequences of class II repeat, two proline stretches each comprising six proline residues were recognized. These proline stretches were separated from each other by Gly and Arg rich 15 amino acids. Two repeats of Pro-Pro-Gly-Arg-Gly-Arg sequence were also present in the proline stretches. Interestingly, the presence of a Arg-X-X-Asp-X-X-Tyr sequence, one phosphorylation site by a tyrosine kinase (24), was recognized in the class II repeat. Moreover, five putative calmodulin-dependent multiprotein kinase sites (R-X-X/S/T)(25), four proline dependent protein kinase sites (X-S/T-P-X)(26), and five protein kinase C sites (S/T-P-R/K)(27) were observed. Three out of five protein kinase C sites were localized near the carboxyl terminus.

The 5'-non coding region. To determine the transcription initiation point of the clone, primer extension analysis was carried out according to Goldberg (20).
Figure 1. Nucleotide sequence of SW2 clone cDNA and predicted amino acid sequence. Nucleotide sequence is numbered from the initiation codon of SW2 protein. Nucleotide sequence of SW1 clone is indicated by a dotted line under the SW2 sequence. Poly A signals are marked with asterisks. Single and double lines show class I and class II repeats of amino acid sequence, respectively. A possible nuclear localization signal, a N-glycosylation and a phosphorylation sites are illustrated by a open box, a gray box and a hatched box, respectively. Proline rich region is indicated by a black box.

out using 5'-end 32P-labeled 84 oligonucleotides located at position -16 to +68. A sharp band could be seen at position -141 on the denaturing gel, indicating that the number of nucleotides extended were 73 when counted from the ATG codon (data not shown). This means that 46 nucleotides are missing in the 5'-non coding region of SW2 clone we obtained when compared with the full length cDNA corresponding to the 2.7 kb mRNA.

The 3'-non coding region. The 3'-non coding region of this clone comprised 1227 nucleotides including two poly A signal sites and 83 bp poly A tail. It should be noted that the difference in length of 3'-non coding regions of both SW2 and SW1 clones was about 760 nucleotides, corresponding to the size difference between the two mRNA species, about 2.7 and 2.0 kb in length. It is also worth mentioning that the sequence of the 3'-non coding region of SW1 clone was completely the same as that of SW2 up to the nucleotide just ahead of the poly A tail of SW1 clone (Fig. 1). It is therefore likely that in SW1 clone, the proximal poly A signal site is functional, while in SW2 clone, the distal poly A signal site is selectively used.

DNA binding properties of SW2 protein

To examine the binding properties of SW2 protein to DNA probe, we prepared its fusion product with E.coli maltose-binding protein (MBP-SW2 protein) and applied to a slot blot competition assay,
Figure 2. DNA binding and competition assay of SW2 protein using slot blot method. 0.3, 1 and 3 μg of purified MBP-SW2 protein and 3 μg of purified MBP-LIF control protein were fixed to nitrocellulose membrane using slot blot. DNA binding activities were determined by probing with 0.1 nanogram of 32P-labeled sense strand, antisense strand and double stranded DNA segment from catalase silencer element. 5000 fold molar excess of oligonucleotides of sense strand (S), 5'-GGTTTGGGGGGAGGG-3'; of antisense strand (AS), 5'-AAACCCCCTCCCCCC-3'; and of nonspecific sequence (SK), 5'-TCTAGAACTAGTGGATC-3' were added as competitors in the binding reaction mixture.

Figure 3. Sequence specific binding of SW2 protein. One μg of purified MBP-SW2 protein was blotted onto a nitrocellulose membrane, and bound to 0.1 nanogram of 32P-labeled antisense oligonucleotide of silencer element in the presence of respective competitors (5000 or 20000 fold molar excess).

Therefore, we propose to refer to it as a dC-stretch binding protein (CSBP).

Binding domain of CSBP (SW2 protein)

Meanwhile, it is worth noting that SW1 was the first clone isolated from the library by using the Southwestern procedure and contained a part of the coding frame comprised of only 84 amino acids residues of the carboxyl terminus. In addition, the binding activity of MBP-SW1 was almost the same as that of MBP-SW2, MBP-SW3, MBP-SW4, while neither MBP-β-gal nor MBP-CNBP used as controls bound to the oligomer probe using 32P-labeled single- and double-stranded probes. Unexpectedly, as little as 0.3 μg of MBP-SW2 protein strongly bound to the dC-rich antisense strand of the silencer sequence. 3 μg of the fusion protein bound slightly to double-stranded DNA probes, but did not to dG-rich sense strands (Fig. 2). Also, no binding to any probes could be seen with unrelated control MBP-LIF. Furthermore, addition of about 5000 fold molar excess of the cold antisense strand probes competed out this binding, while the same amount of a single-stranded sequence (SK sequencing primer for Bluescript II; 5'-TCTAGAACTAGTGGATC-3') did not. Neither binding nor competition could be observed when the dG-rich sense strand was used. The effect of the double-stranded probe was also examined on the same slot blot. It was found that the binding and competition profiles of the double-stranded probe were essentially the same with the antisense probe but to a much lesser extent, indicating a high binding affinity of SW2 protein to the dC-rich single stranded probe.

Sequence specificity for binding with SW2 protein

Next, to determine the size-, the position-, and sequence-specificity of DNA strand for this single-strand binding protein, we prepared various length of dC-rich synthetic oligomers and compared their competitor activities on the slot blot. First, a possible significance of a dT nucleotide between the dC-stretches in the oligomer was examined by applying the following three synthetic oligomers to the competition assay in two doses: C101, C102, C103, containing dA, dG, dC instead of dT, respectively (Fig. 3). All of them, especially C103 oligomer, were found to strongly compete out the binding of the SW2 protein to the probe. We therefore carried out further experiments using oligomers with various length of the dC-stretch. 10mer (C104) and 12mer (C103, C111) of the dC-stretch strongly, but 8mer (C105) weakly bound to MBP-SW2 protein, while 6mer (C106, C109) and 4mer (C107) of the dC failed to show any binding. The binding of the protein to the dC-stretch was weakly retained irrespective of its position in the probe (C105, C108, C110). Furthermore, the dA- and dT-stretches scarcely bound to the protein (C107, C112), while oligomers of cytidine ribonucleotide (prC) bound weakly in the same extent with C105, C108 and C110. No competition activity could be seen in SK oligomers.

These results clearly show that SW2 protein specifically associates to the dC-stretch of the single-stranded DNA.
(Fig. 4). It is therefore evident that the binding site of CSBP to the probe was located within this 84 amino acids. To examine this binding site in more detail, we prepared several 5′- and 3′-deletion mutants of SW1 clone and their fusion proteins. SW11 and SW12 peptides were nine and fifteen amino acids shorter than SW1 at the amino terminus, respectively. Of special interest is that MBP-SW11 protein bound to the probe to the same extent as MBP-SW12 but MBP-SW12 protein completely lost its binding activity. Moreover, we could not detect any binding to MBP-β-galα, MBP-SW4 and MBP-CNBP when the dC-stretch was used as a probe, again confirming the results of the slot blot analysis (data not shown).

These results clearly indicate that 1) CSBP specifically binds to the dC-stretch, and 2) its single strand specific binding domain is located within 75 amino acids residues of the carboxyl terminus.

DISCUSSION

The motivation behind the present study was to isolate a cDNA clone encoding the 35 kDa protein which was previously reported to specifically bind to the G-rich silencer element for catalase gene expression (1). For this purpose, we prepared a double-stranded oligomer by annealing both synthetic sense- and antisense strands of the silencer element and used it as the probe without further purification. It is therefore conceivable that certain amount of the synthetic single-stranded DNA remains without participating in the double strand formation. This may be the reason why the screening of the Xgt11 cDNA library by binding with the probe resulted in the isolation of a cDNA clone encoding a single-stranded DNA binding protein (CSBP) different from our original purpose.

In the present study, we mainly analyzed two clones, SW1 and SW2. The former clone was found to carry a dna fragment encoding a short peptide comprising only 84 amino acid residues of the carboxyl terminus, while the latter was found to carry a near full length cDNA encoding a peptide with 463 amino acid
and cholesterol synthesis (31). RP-A is a 70 kDa subunit of a single-strand binding protein that specifically interacts with oligo-

proteins are related to each other. It is also known that some single-strand binding protein recognize stretches of dC-stretch. and the information about its amino acid sequence is not available. In the present study. H16 protein, however, is not cloned yet. this protein to the dC-rich sequence was inhibited by the addition of dT and involved in the replication of DNA (2,3). The presence of the typical zinc-finger domain in these proteins is worth noting. In our case, however, neither classical zinc-finger motif nor basic-zipper structure, both of which are also often detected in transcription factors could be observed.

ACKNOWLEDGEMENTS
We wish to thank Mr T.Ohmine, Mr K.Adachi and Mrs T.Hara for technical assistance and appreciate to Dr M.Saeki at University of Texas for critical reading of this manuscript. This work was supported by Grants-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan, and by a grant from Uehara Memorial Foundation.

REFERENCES