Nucleic Acids Research, 1992, Vol. 20, No. 23 6317–6321

Determination of recognition-sequences for DNA-binding proteins by a polymerase chain reaction assisted binding site selection method (BSS) using nitrocellulose immobilized DNA binding protein

Peder Lisby Nørby, Niels Pallisgaard, Finn Skou Pedersen and Poul Jørgensen*
Department of Molecular Biology, Aarhus University, C.F.Møllers Alle 130, DK-8000 Århus C, Denmark

Received August 24, 1992; Revised and Accepted November 3, 1992

ABSTRACT

We have developed a simple procedure for rapid determination of a DNA sequence recognized by a DNA binding protein based on immobilization of the protein on nitrocellulose filters. The procedure consists of the following steps: A recombinant protein with a functional DNA binding domain is expressed in *E.coli*. The protein is purified to homogeneity, immobilized on nitrocellulose paper, and exposed to a pool of double stranded oligonucleotides carrying in the central part a 20 bp random sequence, which is flanked by conserved sequences with restriction endonuclease recognition sites for analytical and subcloning purposes and sequences complementary to polymerase chain reaction primers. Oligonucleotides retained by the DNA-binding protein are liberated by increasing the ionic strength and used in a new binding process after amplification by the polymerase chain reaction technique. Finally the amplified product is cloned for determination of the DNA sequence selected by the DNA-binding protein. Murine Zn-finger and basic helix-loop-helix DNA binding proteins were used to demonstrate the efficiency of the method. We show that the yield of oligonucleotides binding to the protein was increased by several consecutive rounds of filter binding and amplification, and that the protein extracted a specific sequence from the pool of random oligonucleotides.

INTRODUCTION

The interaction between DNA and cellular proteins is important in functions that preserve the DNA structure, in DNA repair and replication, and in the process of gene expression. During the characterization of proteins binding to the regulatory regions of genes, it has become clear that the initiation of transcription is an important target for the regulation of transcription and thus for gene expression. By molecular cloning techniques a large panel of genes have been found that express proteins with affinity to the promoter and enhancer regions of a number of genes. Functional studies and comparison of the amino acids of many DNA binding proteins suggested that a limited number of common motifs are responsible for the binding to DNA. Detailed information is available for several DNA binding motifs: zinc finger (1), homeobox domain (2), helix-loop-helix (3), and leucine zipper (4). DNA binding proteins are involved in the permanent alteration in gene expression that takes place during differentiation as first observed in analysis of mutants in Drosophila (2), and in reversible induction of genes as described for the receptor mediated regulation directed by the steroid and thyroid hormones (5,6).

We have cloned a number of murine DNA binding proteins in a modified λgt11 (λgt11*Sfi-Not*) (7) using the filter-binding method described by Singh *et al* (8) and by Vinson *et al* (9). Based on previously described methods (10–15), we designed the PCR assisted binding site selection method (BSS) using DNA binding protein immobilized on nitrocellulose paper to determine the DNA sequence recognized by these DNA binding proteins. The process requires a protein, where the DNA-binding capability is preserved. In our analysis, we have used *E.coli* produced chimerical proteins expressed from pGEMEX (Promega) and thus fused to a segment of bacteriophage T7gene*10* product as well as intact proteins expressed from pET3ASEN (16). The purified proteins were immobilized on nitrocellulose filters and exposed to a double stranded oligonucleotide (ds-oligo), which carried a twenty bp random sequence that was flanked by preserved sequences with restriction endonuclease recognition sites for analytical and subcloning purposes and with ends complementary to two PCR primers. Using molecularly cloned murine proteins carrying either a basic helix-loop-helix (ALF1 and ALF2) or a Zn-finger DNA binding motif (ZEB), we demonstrate that each of these proteins are capable of extracting a specific DNA sequence from the pool of random ds-oligo's.

* To whom correspondence should be addressed

MATERIAL AND METHODS

Expression and purification of DNA binding proteins

The expression vector pGEMEX2 was obtained from Promega and pET3ASEN was derived from pET3A (17,18) by deletion of the *EcoRI* site and insertion of a *NdeI-SfiI-EcoRI-NotI* polylinker, in frame with the *lacZ* gene of λgt11*Sfi-Not*, between the *NdeI* and *BamHI* sites (16). Plasmids expressing the DNA binding proteins were constructed by subcloning of the cDNA from λgt11*Sfi-Not* either directly as a *SfiI/NotI* fragment or as an *EcoRI/NotI* fragment via pBluescript SK (Stratagene). The DNA binding domain of the ZEB cDNA (unpublished) was expressed in pET3ASEN and had the molecular weight of 38 kD, whereas the DNA-binding domains of ALF1 and ALF2 (16) were expressed in pGEMEX2 as T7gene*10* fusion proteins with molecular weight of 79 and 70 kD, respectively. The constructs were transformed into *E.coli* BL21(DE3)pLysS (18). Exponentially growing bacteria at 37°C in M9ZB medium (18) was induced at OD_{600} = 0.5 with IPTG (0.5 mM) for 2.5 h before harvest. The protein was purified to near homogeneity as inclusion bodies by a method modified from Bohmann and Tjian (19): Bacteria from a 200 ml culture were harvested by 10 min centrifugation at 5000 rpm in a Sorvall GSA rotor. The pellet was resuspended in 33 ml of 0°C buffer A (50 mM Tris−HCl [pH 8.0]; 5 mM EDTA; 0.1% v/v Triton X-100; 1 mM DTT; 1mM PMSF), sonicated 5 times for 15 sec in a MSE sonicater, using a large probe at setting 18. The inclusion bodies were sedimented by a 10 min centrifugation at 0°C and 15000 rpm in a Sorvall SS34 rotor. The pellet was suspended by a short pulse of sonication at 0°C in 33 ml buffer B (10 mM Tris−HCl [pH 8.0]; 1 mM EDTA; 0.5 M LiCl; 0.5 % v/v dodecyldimethylamine (Serva); 1 mM DTT). The suspension was centrifuged for 10 min at 0°C and 15000 rpm. The pellet was washed by a sonication aided resuspension in Buffer B and twice with Buffer C (10 mM Tris−HCl [pH 8.0]; 1 mM EDTA; 0.5% v/v dodecyldimethylamine; 1 mM DTT) and centrifuged as before. The inclusion bodies were solubilized at 0°C in 8 ml Buffer D (25 mM Hepes [pH 7.9]; 10 mM $MgCl_2$; 20% v/v glycerol; 40 mM KCl; 0.1% v/v dodecyldimethylamine; 1 mM DTT) with 5 M guanidine-HCl and incubated overnight at 4°C with gentle shaking. The proteins were renatured by dialyses for periods of 2 h at 4°C against buffer D containing 2 M, 1 M, 0.5 M, 0 M, 0 M, and 0 M guanidine-HCl, respectively, and stored at −80°C. The yields were in the order of 10 μg protein per ml bacterial culture.

Construction of oligonucleotides

The random oligonucleotide and the PCR primers , shown in Fig. 1, were synthesized on a 380A Applied Biosystem DNA synthesizer. A template oligonucleotide (BSS-1) was designed to contain a randomized sequence of twenty bases. The random sequence was flanked by recognition sites for *XbaI* and *BamHI* in the conserved sequences to be used for subcloning purposes. A *BglII* was included for the identification of plasmids with insert. Furthermore two primers (BSS-2 and BSS-3) complementary to the conserved regions on the random primer were designed for the PCR amplification. The probe used to set up the filter binding conditions was a ds-oligo complementary to a segment in the Akv murine leukemia virus enhancer region carrying the E-box motif (20).

Immobilization of DNA binding proteins on nitrocellulose filters

The concentration of the purified protein was estimated to 100 μg per ml by visual inspection of a SDS-PAGE gel. The purified protein (2 μl) was spotted onto a 1 cm^2 square of nitrocellulose filter. After 10 min of air drying the filter was blocked for 30 min at 4°C in binding buffer (25 mM Hepes [pH 7.9], 40 mM KCl, 3 mM $MgCl_2$, 1 mM DTT) containing 0,5% carnation milk. Filters were washed in binding buffer with 0.25% carnation milk for 15 min before exposure to the random ds-oligo mixture.

The primary binding of the random ds-oligo BSS-1

Before the first binding the random oligo was converted to double stranded DNA by DNA polymerase I Klenow fragment extension from the BSS-3 primer annealed to the 3′ end of BSS-1. The DNA binding protein immobilized on nitrocellulose filters was exposed to roughly 8 pmol of the double stranded random oligonucleotide mixture for 2 h or overnight at 4°C in 200 μl binding buffer. After oligonucleotide binding the nitrocellulose squares were washed 3 times for 5 min in binding buffer with 0.25% carnation milk and once in binding buffer, all steps at 4°C.

The binding site selection cycle

The bound ds-oligo was dissociated from the DNA binding protein by a wash in 200 μl 0.5 or 1.0 M KCl for 10 min at 4°C. For PCR amplification 10 μl or 5 μl of the rescued ds-oligo in 0.5 or 1.0 M KCl, respectively, was mixed with 10 μl 10×PCR buffer without KCl (100 mM tris-HCl; 15 mM $MgCl_2$ [pH 8.3]), 10 μl 2 mM dNTP, 10 μl BSS2 primer (10 μM), 10 μl BSS3, and 2.5 units Taq polymerase (Stratagene) in a final volume of 100 μl. Twenty cycles of amplification were done in a Hybaid OmniGene thermocycler with denaturation at 95°C for 50 sec, annealing at 64°C for 50 sec, and extension at 72°C for 1 min using the tube temperature control mode. Five μl of the amplified ds-oligo was used for a new round of filter binding without any steps of purification.

Cloning and sequencing of the oligonucleotide selected by the BSS method

After the final PCR step the pool of oligonucleotides was digested with *XbaI* and *BamHI* and inserted into the corresponding sites in pBluescript KS. Recombinant plasmids were detected as white

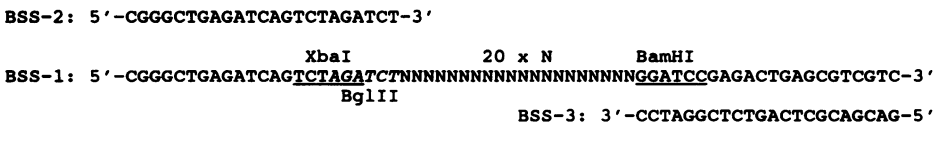

Figure 1. Three oligonucleotides used in the PCR assisted binding site selection method. BSS-1 was designed to include random oligonucleotides at twenty positions flanked by specific sequences at each end and recognition sequences for the restriction endonucleases *XbaI* and *BamHI* for subcloning purposes and *BglII* to be used in the analysis of the plasmid subclones. BSS-2 and BSS-3 are primers employed for generation of double stranded oligonucleotides and for PCR.

colonies on IPTG-XGal-Amp plates. The presence of insert was verified by digestion with *Bgl*II of mini-prep plasmid DNA. The DNA sequences were determined using an Applied Biosystem sequencing kit on an Applied Biosystem 373A sequencing apparatus.

RESULTS

Selection of the binding sequence for DNA binding proteins from a pool of random double stranded oligonucleotides

Analysis of the mechanisms of transcription control in eukaryotic cells has revealed the importance of interaction between cellular proteins and the regulatory regions of the genes. The characterization of DNA-binding proteins has been made by conventional biochemical methods, and more recently by direct molecular cloning using the method of Singh *et al* (8) and Vinson *et al* (9). This method is based on expression of the cDNA segment from a λ vector in *E.coli*, transfer of the expressed recombinant protein to nitrocellulose paper, and detection by a double stranded DNA probe complementary to the potential target sequence. By these methods we have, from the murine fibroblast cell line NIH3T3, identified a number of λgt11*Sfi-Not* cDNA clones that code for proteins with affinity to the enhancer-promoter region of the murine leukemia virus Akv. DNA binding proteins differ greatly in their affinity to DNA, some proteins have general affinity to DNA sequences, whereas others have a very specific recognition sequence in the DNA. In our cloning protocol the expressed proteins were detected by a specific sequence, the enhancer-promoter sequence of Akv murine leukemia virus. To examine the sequence specificity of these DNA binding proteins, we have used a PCR assisted binding site selection (BSS) method shown schematically in Figure 2, where the DNA-binding protein was immobilized on a nitrocellulose membrane and allowed to select its binding sequence from a pool of random double stranded oligonucleotides, shown in Figure 1. The bound oligonucelotides were liberated by increasing the ionic strength, amplified by PCR, and cloned for determination of their DNA sequence.

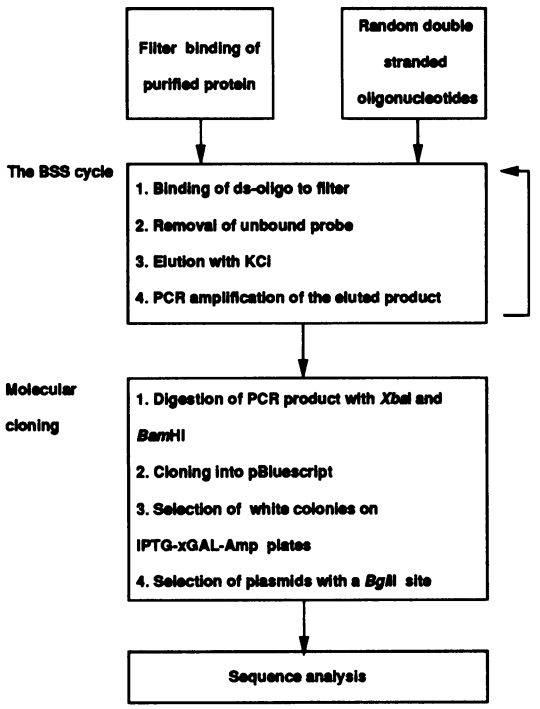

Figure 2. Flow scheme for PCR assisted binding site selection (BSS): The binding site selection procedure (BSS) requires a pool of random double stranded oligonucleotides and a DNA binding peptide either as a fusion protein, a partial, or an intact protein.

Parameters in the binding and dissociation DNA protein interactions

There are three critical steps in the BSS method: (i) the binding of the oligonucleotide to the protein, (ii) the removal of the non-bound oligonucleotides, and finally (iii) the recovery of the oligonucleotides bound to the DNA binding protein. The DNA-binding proteins form a large group and they vary greatly in their DNA binding activity. We used the ionic strength to direct the binding and dissociation of the double stranded oligonucleotides to the nitrocellulose immobilized DNA-binding protein. When the λ clones expressing the DNA-binding protein is found by the method of Vinson *et al* (9) at least one oligonucleotide is available to monitor the binding parameters. The proteins ALF1, ALF2, and ZEB used as examples here bind to an oligonucleotide complementary to a sequence in the enhancer region of Akv murine leukemia virus (7,16, and unpublished observations). We used this oligonucleotide to characterize the binding-dissociation conditions used in the BSS cycles, and the abundance in the PCR

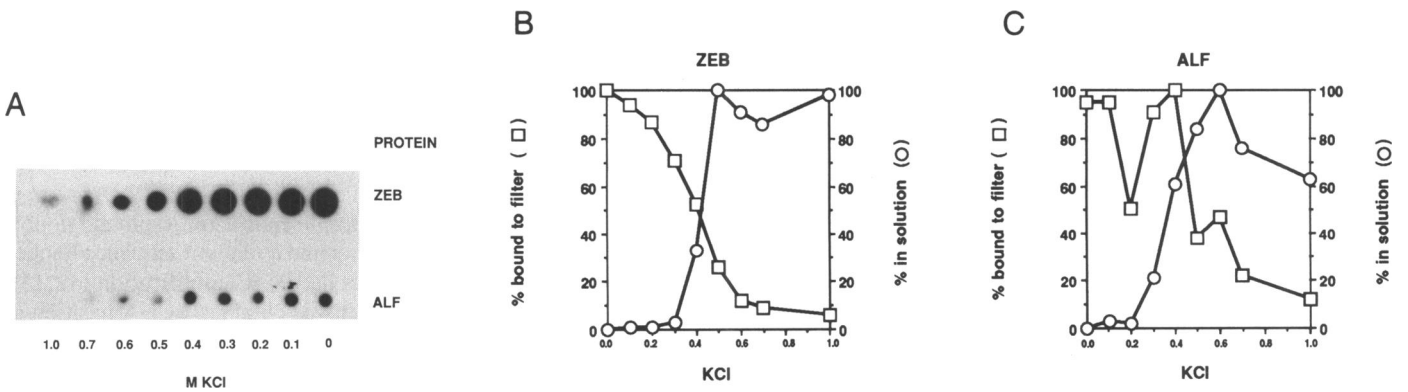

Figure 3. Elution of a radioactive probes from nitrocellulose immobilized DNA-binding proteins. The recombinant proteins ZEB and ALF were immobilized on nitrocellulose filters. Radioactively labelled probe was bound to the filters and the filters were washed with increasing concentration of KCl as indicated. **A**, residual radioactivity was monitored by radioautography. **B** and **C**, radioactivity on the filter and in solution were estimated by scintillation counting (B, ZEB; C, ALF).

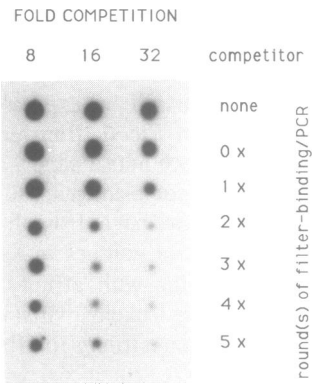

Figure 4. Estimation of the yield of double stranded oligonucleotides with affinity to ALF proteins at the various steps of BSS. The recombinant protein ALF was immobilized on nitrocellulose filters. The PCR product at three concentrations were bound to the protein in presence of a radioactively labelled probe as described in Material and Methods. None, without competitor; 0 ×, competition with the ds-oligo mixture used in the first filter binding; n ×, competition with PCR product after n rounds of filter binding(s).

amplified pool of oligonucleotides with affinity to the protein by a competition assay. The primary association of the oligonucleotide to the DNA-binding protein was measured by binding a radioactively labelled probe to the DNA-binding protein, when immobilized on nitrocellulose filter (data not shown). To elaborate procedures to recover ds-oligo retained by the DNA-binding protein on the filter, we tested guanidine-HCl, NaOH, and KCl and found that all dissociated the protein-oligo complex (data not shown). Because KCl is less disturbing in the subsequent steps, we chose this to liberate the ds-oligo from the DNA-binding protein and determined the concentration necessary for the dissociation, see Fig. 3. The high salt concentration does not interfere with the subsequent steps because the DNA is diluted before PCR amplification. For the ALF and ZEB proteins 0.4 M KCl was sufficient to remove 50% of the bound probe. Based on these results, we decided to use a concentration of KCl above 0.5 M to recover the bound ds-oligo from the filter. To monitor the relative yield of sequence with specific affinity to the DNA-binding protein at the various steps of the BSS procedure, we utilized the PCR product to compete a radioactive oligonucleotide probe with specific affinity to the DNA-binding protein in a filter binding assay. The procedure is illustrated in Figure 4, which shows that two rounds of filter binding-PCR are sufficient to yield a significant concentration of oligonucleotides with affinity to the ALF protein. Thus the frequency of plasmids with the consensus DNA sequences recognized by the DNA-binding protein may be increased by several consecutive rounds of filter binding and PCR amplification. It should be emphasized that the optimal binding conditions vary for individual DNA-binding proteins.

Cloning and sequencing of the BSS double stranded oligonucleotide

After five rounds of filter binding/amplification the *Xba*I-*Bam*HI fragments were recovered from the pool of oligonucleotides and cloned into pBluescript KS for sequence analysis. From the ALF2 experiment, thirteen clones that carried the *Bgl*II site from the random oligonucleotide were sequenced. ALF2 is a basic helix-loop-helix protein proposed to have affinity to an E-box sequence in the murine leukemia virus enhancer sequence (CAGATG)

```
BSS#75       ggatccACTAATCCGCGCAGGTGTTCagatctaga
BSS#76               ggatccGCAGGTGCAANCAGGGTGTGagatctaga
BSS#77      ggatccGGCTGCANCAGACAGGTGTGagatctaga
BSS#78            tctagatctAACAGCTGTTAACAGGGCGGggatcc
BSS#79     tctagatctGGTCAGGTGACAGGTGCGTggatcc
BSS#80            tctagatctGCAGGTGGCCGCTAGCGGCCggatcc
BSS#81         tctagatctGATTACAGCTGATGGCAGGTggatcc
BSS#82  tctagatctCAGGTGGCCTGTACAGGTGCggatcc
BSS#84           tctagatctTAGCAGGTGCATAATCAGGCggatcc
BSS#86           tctagatctCGGCAGGTGCTTGATCAGGTggatcc
BSS#88             tctagatctGCAGGTGTATTGAGGGGGGggatcc
BSS#90          tctagatctGGAACAGGTGGTCCAGGTGTggatcc
BSS#91      tctagatctCACGANGNGCAGGTGTNGCTggatcc

ALF2 consensus:              ACAGGTGC
                             G   C  T

Akv E-box:       ..CAGGGCCAAGAACAGATGGTCCCCA..

E2A (E12/E47) consensus:   NNGCAGGTGTNN
                            A   C  G
                                A
```

Figure 5. Alignment of the ALF2 binding sites found by the BSS method. The PCR amplified double stranded oligonucleotides after five rounds of BSS were subcloned between the *Xba*I and *Bam*HI sites in pBluescript KS. DNA sequences of the inserts in randomly picked plasmid carrying the *Bgl*II recognition site were determined. BSS#, refer to individual clones; ALF2 consensus, consensus recognition sequences of ALF2; Akv E-box, the E-box sequence of Akv murine leukemia virus (16); E2A (E12/E47) consensus, consensus recognition sequences of the human E12/E47 proteins (12). Underlined sequences, E-boxes; lower-case letters, linker sequences.

(16), and is the murine homologue to the human protein E47 (3). The E47 consensus recognition sequence has been determined by a band shift method (12). Using the BSS method ALF2 extracted the same consensus sequence (see Figure 5). In BSS analysis the ALF1 and ZEB proteins yielded consensus sequences that were compatible with the Akv E-box (CAGATG). The ALF1 consensus binding sequence was identical to the canonical E-box sequence (CANNTG), whereas ZEB recognized the consensus sequence NNGATG, and thus included the central GA of the Akv E-box in its target sequence (data not shown). We noted for all three proteins several clones with multiple recognition motifs in the inserted sequence.

DISCUSSION

A simple and multifaceted method is presented, called binding site selection (BSS), that allows determination of target sequences for putative DNA binding proteins. The BSS method represents an improvement of previously described procedures for determination of the target sequence of DNA and RNA binding proteins (10−15). In these a purified DNA-binding protein is typically allowed to select its target sequence from a mixture of random oligonucleotides, and the final step in the procedures is a molecular cloning and a determination of the selected target sequence in some cases after an amplification by PCR. As a target for the DNA-binding protein has been utilized synthetic double stranded oligonucleotides that contain a random sequence flanked by conserved sequence used in the PCR amplification (12,13), or random DNA fragments that was obtained by a sonication of genomic DNA and supplemented with oligonucleotide linkers subsequently employed in the PCR amplification (11). The DNA−protein complex has been recovered by band-shift (12) or by filtration through a nitrocellulose filter (13). Alternative a DNA binding protein may be bound to a sepharose column, which is then used for the extraction of the target sequence from

the pool of random oligonucleotides (10). Due to the capacity of the column this method yields sufficient material for a direct cloning without PCR amplification.

We have used purified recombinant DNA-binding protein immobilized on nitrocellulose paper to extract the target sequence from the pool of random sequences. The BSS method was developed to be used in the primary characterization of λ cDNA clones which expressed DNA binding proteins. It requires a small amount of purified protein including the DNA binding domain. For this purpose we have subcloned the protein in *E.coli* plasmid expression vectors pGEMEX (Promega) or pETASEN (16). It should be emphasized that only the segment coding for the DNA binding domain is essential for the determination of the target of the DNA binding protein. The BSS may supplement the above mentioned methods in the determination of the *in vitro* binding of purified DNA binding proteins. The target DNA sequence of the protein complexes existing *in vivo* may be recovered by immunoprecipitation of the DNA-protein complex using antibodies against the DNA binding protein from nuclear extracts supplemented with a random probe (14). The ultimate result of the BSS procedure is the determination of the consensus binding sequence after subcloning of the selected pool of oligonucleotides. The affinity to the DNA-binding protein of the individual cloned target sequence may be further tested by band-shift analysis (12,13) or their function *in vivo* may be tested by subcloning the target sequence into a reporter plasmid and co-transfecting it with a plasmid expressing the DNA-binding protein.

We demonstrate the BSS-method on one of our basic helix-loop-helix proteins, termed ALF2 (16), which is the murine homologue to the human DNA-binding protein E47 (3). Using the BSS method, we show that ALF2 and E47 recognized identical target sequences. As the E47 target sequence was extracted from a pool of random oligonucleotides by the band shift method (12), we conclude that the BSS technique functions as well as the method based on the band-shift technique. Moreover extraction of the target sequence from the pool of random ds-oligo's by nitrocellulose immobilized proteins makes the BSS easier to use than previously described procedures. The major advantage of the BSS method is that it is extremely simple. This permits a simultaneous analysis of a large number of individual DNA-binding proteins, which is the case during the primary cloning of genes for DNA-binding proteins or in analysis of a gene for a specific DNA-binding protein by introduction of mutations. We trust that the BSS will supplement the previously described methods for target determination of DNA binding proteins.

ACKNOWLEDGEMENTS

We thank Kirsten Paludan for critical reading of the manuscript. This project was supported by The Danish Biotechnology Program and the NOVO foundation.

REFERENCES

1. Miller,J., McLachlan,A.D. and Klug,A. (1985) EMBO J, 4, 1609–1614.
2. Levine,M. and Hoey,T. (1988) Cell, **55**, 537–540.
3. Murre,C., McCaw,P.S. and Baltimore,D. (1989) Cell., **56**, 777–783.
4. Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) Science., **240**, 1759–1764.
5. Yamamoto,K.R. (1985) Annu. Rev. Genet., **19**, 209–252.
6. Evans,R.M. (1988) Science, **240**, 889–895.
7. Pallisgaard,N., Pedersen,F.S., Kjeldgaard,N.O. and Jørgensen,P. (1990) Gene Regulation and AIDS: Transcriptional activation, retroviruses, and pathogenesis **7**, 87–94.
8. Singh,H., LeBowitz,J.H., Baldwin,A.S.J. and Sharp,P.A. (1988) Cell, **52**, 415–423.
9. Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) Genes Dev., **2**, 801–806.
10. Oliphant,A.R., Brandl,C.J. and Struhl,K. (1989) Mol. Cell. Biol., **9**, 2944–2949.
11. Kinzler,K.W. and Vogelstein,B. (1990) Mol. Cell. Biol., **10**, 634–642.
12. Sun,X.H. and Baltimore,D. (1991) Cell., **64**, 459–470.
13. Thiesen,H. and Bach,C. (1990) Nucleic Acids Res., **18**, 3203–3209.
14. Pollock,R. and Treisman,R. (1990) Nucleic. Acids. Res., **18**, 6197–6204.
15. Tuerk,C. and Gold,L. (1990) Science, **249**, 505–510.
16. Nielsen,A.L., Pallisgaard,N., Pedersen,F.S. and Jørgensen,P. (1992) Mol. Cell. Biol., **12**, 3449–3459.
17. Studier,F.W. and Moffatt,B.A. (1986) J. Mol. Biol., **189**, 113–130.
18. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Methods. Enzymol., **185**, 60–89.
19. Bohmann,D. and Tjian,R. (1989) Cell., **59**, 709–717.
20. Ephrussi,A., Church,G.M., Tonegawa,S. and Gilbert,W. (1985) Science., **227**, 134–140.