*Proc. Natl. Acad. Sci. USA*
Vol. 93, pp. 1184–1189, February 1996
Biochemistry

# Multiplex selection technique (MuST): An approach to clone transcription factor binding sites

**(transcription factors/T cells/cold shock proteins/Ets protein)**

GIRISH N. NALLUR, KULKARNI PRAKASH, AND SHERMAN M. WEISSMAN*

Department of Genetics, Boyer Center for Molecular Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06511

*Contributed by Sherman M. Weissman, October 20, 1995*

**ABSTRACT** **We have used a multiplex selection approach to construct a library of DNA–protein interaction sites recognized by many of the DNA-binding proteins present in a cell type. An estimated minimum of two-thirds of the binding sites present in a library prepared from activated Jurkat T cells represent authentic transcription factor binding sites. We used the library for isolation of "optimal" binding site probes that facilitated cloning of a factor and to identify binding activities induced within 2 hr of activation of Jurkat cells. Since a large fraction of the oligonucleotides obtained appear to represent "optimal" binding sites for sequence-specific DNA-binding proteins, it is feasible to construct a catalog of consensus binding sites for DNA-binding proteins in a given cell type. Qualitative and quantitative comparisons of the catalogs of binding site sequences from various cell types could provide valuable insights into the process of differentiation acting at the level of transcriptional control.**

The importance of transcription factor cascades in cell proliferation and differentiation is a major focus of study in modern biology. Studies of transcriptional control mechanisms underlying spatially restricted transcription in the early embryonic development of *Drosophila* (1), lineage specification in muscle (2) and nerve cells (3), mammary cell differentiation (4), and several other examples illustrate that gene–protein interactions initiated by key transcription factors bring about long-term changes in gene expression, often involving the activation of other transcription factors that control downstream processes. Cellular differentiation and proliferation can be regarded, therefore, as an integrated process involving the concerted and sequential action of transcription factors that determine the biology of that cell type.

A typical cell in humans expresses about 10,000 genes. By extrapolation from lower organisms, about 2–20% of these genes may encode transcription factors; each cell type could thus express as many as 200–2000 different transcription factors. Identifying transcription factors has been a time-consuming and tedious task that often requires large amounts of biological material or depends upon the availability of mutants and, in general, has yielded one new factor per investigation. We have exploited the sequence-specific DNA binding properties of transcription factors in an approach for direct isolation of binding sites recognized by a large number of protein factors. These binding sites serve as efficient probes for isolation of the cognate factors. The approach is rapid and can be reiterated to derive progressively more information from the products of each study.

In this protocol, a library of binding sites is constructed by repeated selection of a random pool of oligonucleotides by electrophoretic mobility-shift assay (EMSA) with nuclear extracts under conditions that favor selection of preferred binding sites. Activated Jurkat cells were used as a model system for identification and cloning of DNA sites that are specifically recognized by protein factors. We used the library probes for isolation of "optimal" sites for cloning specific transcription factors and for probing qualitative differences in transcription factor expression between resting and activated T cells. Potential applications of the approach to identify and clone transcription factors specific to a cell type are discussed.

## MATERIALS AND METHODS

**Cells.** Jurkat cells were grown in RPMI 1640 medium containing 10% (vol/vol) fetal bovine serum and stimulated with phorbol 12-myristate 13-acetate and phytohemagglutinin as described (5).

**EMSA.** EMSA was done as described (5) with a few modifications. A typical binding reaction carried out for 10 min on ice contained 1 μl of nuclear extract (1.4 μg of protein) and 1–2 ng of $^{32}$P-labeled probe in a 10-μl reaction volume containing 1 μg of poly(dI-dC). Protein–DNA complexes were fractionated on a 5% polyacrylamide gel in 0.5× TBE (1× = 90 mM Tris/64.6 mM borate/2.5 mM EDTA, pH 8.3). Competitions were done with a 40- to 50-fold excess over the probe amounts. Extraction and PCR amplification of oligonucleotide probes from the EMSA gel were performed as described (6). Crude nuclear protein extracts were prepared from resting (HA-R) and activated (HA-TP2) Jurkat cells as described (7) and fractionated on a heparin-agarose column (8). Synthetic oligonucleotides used for the selection rounds contained the sequence 5′-CGAGGTCGACGGTATCGNNNNNNNNNN-GGATCCACTAGTTCTAGAGC-3′. PCR primers had the following sequences: P1, 5′-CGAGGTCGACGGTATCG-3′; and P2, 5′-GCTCTAGAACTAGTGGATC-3′. Four rounds of reiterative selections were performed by EMSA with these oligonucleotides and HA-TP2 extract. The selected oligonucleotides, TP2-4, were cloned in pBluescript vector to obtain a library of binding sites (TP2-4/BS). The sequences of synthetic oligonucleotides used in this study are as follows: BSO8, 5′-GAATAGAACTGGATC-3′, BSO3/9, 5′-ATCGG-TAGG(G or T)G(G or C)GTCGA-3′; S6-1, 5′-ACCCCA(T or C)ACT-3′; S6-2, 5′-CCATGATTAC(G or A)-3′; and S6-3, 5′-ACCGGAAG(T or C)T-3′. The S6-1, -2, and -3 oligonucleotides contained the heel sequence 5′-CTTTAGTGAGG-GTTAAT-3′ at the 3′ end. Radioactive probes for EMSA were prepared either by extension of annealed primers with Klenow fragment polymerase or were end-labeled with phage T4 polynucleotide kinase in the presence of $^{32}$P-labeled nucleotides. The wild-type octamer (wt-Oct) (9) and nuclear transcription factor NF-AT (10) probes used in these studies have been described.

For analysis of gel segments, an EMSA experiment was performed with the TP2-4 probes and HA-TP2 extract. This lane was divided into six segments (of 1 cm each) correspond-

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: EMSA, electrophoretic mobility-shift assay; wt-Oct, wild-type octamer; MuST, multiplex selection technique.

*To whom reprint requests should be addressed.

ing to the upper third of the gel. Oligonucleotides contained in each segment were separately eluted, amplified, and used for EMSA analysis.

## RESULTS

We applied the principles of multiplex selections to construct a library of binding sites from activated Jurkat cells. Briefly, synthetic oligonucleotides randomized at 10 contiguous positions and flanked by PCR heel sequences, P1 and P2, were used as probes for EMSA with nuclear extracts. The sequences of P1 and P2 are described in *Materials and Methods*. The portion of the EMSA gel containing bound complexes was excised, and the oligonucleotides were eluted and amplified by PCR. An aliquot of the amplified product was used as the probe for the next cycle of selection by EMSA. After four rounds of selection, the amplified oligonucleotides (TP2-4) were cloned to obtain a library of binding sites (TP2-4/BS).

Initially, to evaluate the method, we obtained sequences of 20 clones randomly picked from this library. Among these, 8 binding site sequences resembled those for known transcription factors (Table 1; refs. 11–17). Oligonucleotides containing sequences of clones for 3 known and 4 unknown sites were analyzed by EMSA. Of these, 4 oligonucleotides exhibited specific shifts (Fig. 1). By extrapolation, more than two-thirds of the binding sites contained in the library represent authentic binding sites for DNA-binding proteins.

To ascertain that bands extracted from any region of the gel and used as probes in a subsequent EMSA experiment regenerate a band migrating to the original position, we performed an EMSA experiment with the HA-TP2 extract and TP2-4 probes (not shown). Oligonucleotides contained in various segments were extracted and used as probes for analysis by EMSA. Each probe produced a competable band at the corresponding position (Fig. 2, lanes 1–6). Most of the segment probes showed additional competable bands, revealing the presence of other proteins or protein complexes that bind to the site(s) contained in that gel segment.

Table 1. Sequences of a sample of clones from the binding site library

| Name | Binding site | Factor | EMSA |
|---|---|---|---|
| BSN4 | ggaCCAATcg | CCAAT-bf (11) | ND |
| BSN7 | TGGGtGGGGT | SP1 (12) | ND |
| BSN8 | accTGATAtg | GATA-1 (13) | ND |
| BSN12 | gTGACTCccg | TPA Resp Ele (14) | ND |
| BSN20 | gcCAAGGTCg | ELP (15) | ND |
| BSN15 | CTTTA_CTcG | ANF (16) | Y |
| BSN5-1 | cCcATCCATg | Pit-1 (17) | N |
| BSN17 | GTAgCCATgt | Pit-1 (17) | Y |
| BSN3 | ATGGTAGCGT | Unknown | N |
| BSN5-2 | TCGCAATGGG | Unknown | Y |
| BSN11 | GGTAATAGGN | Unknown | Y |
| BSN18-3 | CTGGATTATG | Unknown | N |
| BSN1 | GTGGGGGGTT | Unknown | ND |
| BSN6 | GTGGGAGGGT | Unknown | ND |
| BSN9 | CATGGCGGAC | Unknown | ND |
| BSN12 | GTGACTCCCG | Unknown | ND |
| BSN13 | GTCTAGGGAG | Unknown | ND |
| BSN14 | GGGGGGCTAT | Unknown | ND |
| BSN18-1 | GGGTTATAG | Unknown | ND |
| BSN18-2 | GATTTGGACT | Unknown | ND |

The sequences of 20 clones from the TP2-4/BS library were obtained and compared with sequences present in the transcription factor data base (National Institutes of Health). Similarities with previously known transcription factor binding sites are shown. Those that did not match with sequences in the data base have been designated as unknown. Results from the EMSA analysis (see Fig. 1) are summarized. Sequences marked ND were not analyzed by EMSA; Y = positive shift; N = no shift. Numbers in parentheses indicate references.

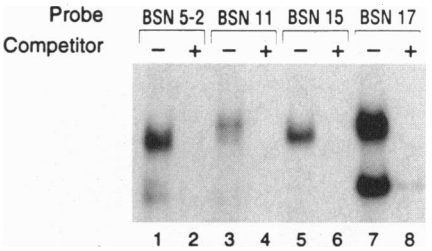


FIG. 1. EMSA analysis of randomly picked clones. Synthetic oligonucleotides for seven of the randomly picked clones (see Table 1) were used as probes for EMSA with the HA-TP2 nuclear extract. Those that exhibited positive EMSA shifts are shown. Lanes marked "competitor +" represent EMSA performed in the presence of a 50-fold excess of unlabeled probe.

Several factors present in T cells have been shown to complex with the wt-Oct sequence (5, 18, 19). We confirmed the presence of octamer (Oct) binding sites in our library, using bacterially produced Oct2 protein (Fig. 3*A*, lane 10). However, an EMSA with a synthetic wt-Oct probe and HA-TP2 extract revealed several competable bands (Fig. 3*A*, lanes 1 and 2). In particular, a prominent band (OctT3) has been observed (5, 20) as a protein antigenically distinct from either Oct1 or Oct2. However, attempts in our laboratory to clone OctT3 cDNA(s) by screening expression libraries in bacteria with a multimerized wt-Oct probe were unsuccessful (K.P. and S.M.W., unpublished data). Therefore, we took an alternate approach to clone cDNA(s) corresponding to the OctT3 EMSA band.

It was likely that the OctT3 activity associated fortuitously with the wt-Oct probe and that other preferred sites for OctT3 existed. If this were so, then such sites should have been selected during the random oligonucleotide selections with total nuclear extract and should be represented in the library of binding sites. Therefore, we tried to isolate high-affinity binding sites for OctT3 from the TP2-4 probes by competition of the OctT3–wt-Oct complex. The resulting probe mixture strongly competed with the OctT3 band and also with bands 4 and 6 in a subsequent EMSA with the wt-Oct probe (data not shown). Upon cloning and sequencing nine clones from this probe mixture, five clones contained the A+T-rich sequence 5′-AATAGAA-3′ (BSO8), while three others were a set of G+C-rich sequences that showed the consensus pattern 5′-TAGGDGBAGGG-3′ (BSO3/9) (Fig. 3*B*). The BSO8 probe specifically competed with the OctT3 band and with bands 4

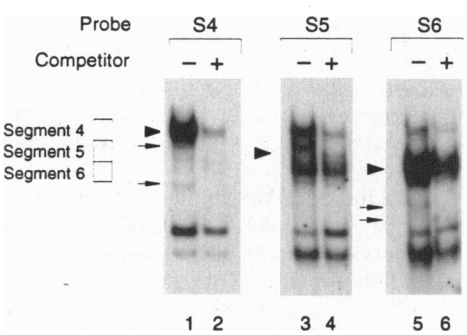


FIG. 2. Analysis of gel segments shows reproducibility of band patterns and reduces complexity. An EMSA experiment was performed with the TP2-4 probes and HA-TP2 extract (not shown). Shifted oligonucleotides present in a 1-cm length of the lane were isolated and used as probes for EMSA. Segments 4, 5, and 6 (represented by squares) indicate the gel region from which the EMSA probes were extracted. The expected complexes are marked by arrowheads. Some additional complexes are marked by thin arrows. Competitions were as in Fig. 1.

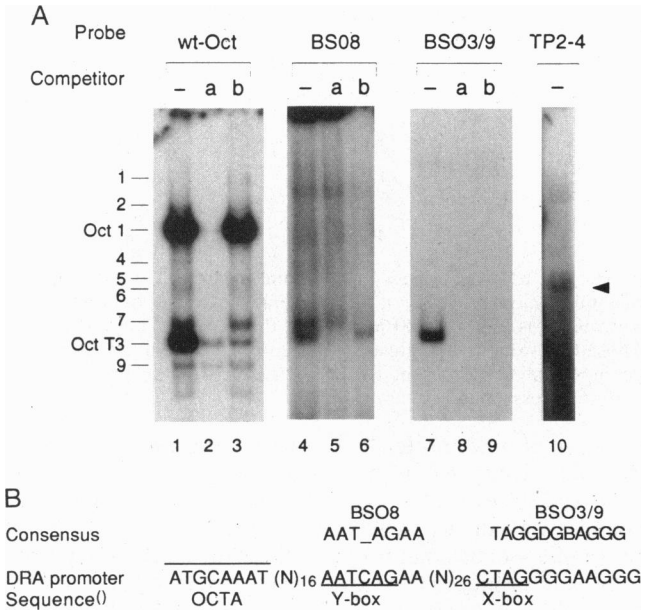


FIG. 3. Isolation of optimal binding sites for OctT3. (*A*) An EMSA experiment was performed with a synthetic wt-Oct duplex probe and HA-TP2 extract in the presence of competing TP2-4 probes. The region of the EMSA gel corresponding to the OctT3 band position was excised, amplified by PCR, and sequenced. Synthetic probes representing the consensus sequences—BSO8 and BSO3/9 (see *Materials and Methods*)—were used in EMSA experiments with the HA-TP2 extract and for reciprocal competitions with the wt-Oct probe. Competitor oligonucleotides were wt-Oct (lanes a) and BSO8 (lanes b). A single complex obtained in an EMSA experiment with bacterially produced Oct2 protein and TP2-4 probes is marked by an arrowhead (lane 10) (see *Results*). (*B*) Comparison of BSO8 and BSO3/9 consensus sequences with the major histocompatibility complex class II *HLA-DRA* promoter sequence (bottom strand). The X- and Y-box sequences are underlined. The consensus wt-Oct sequence is overlined. The number of intervening nucleotides between these boxes is shown in parentheses.

and 6 that had been seen with the wt-Oct probe (Fig. 3*A*, lane 3). Several complexes were seen when BSO8 itself was used as probe (lane 4) that were competitively blocked by excess BSO8 (lane 6) but only partially by the wt-Oct probe (lane 5). The BSO3/9 EMSA pattern consisted of a single prominent band in the position of a major band of the BSO8 pattern (Fig. 3*A*, compare lanes 4 and 7). Nevertheless, BSO3/9 binding was competitively blocked by the BSO8 probe as well as the wt-Oct duplex (Fig. 3*A*, lanes 8 and 9, respectively), suggesting that factors that bind to the BSO3/9 probe could be a subset of those that bind to the BSO8 probe. Taken together, these results showed that the BSO8 binding site was selected from the library in preference to the wt-Oct sequence. This finding strongly suggested that BSO8 is an optimal binding site for the OctT3 protein.

We screened a Jurkat cDNA library in phage λgt11 (21) with the BSO8 probe. We isolated four cDNAs, all of which encoded members of a family of transcription factors termed the Y-box-binding proteins. One of the clones, 30b3 was identical to YB-1 (22). Two other clones were presumptive splice variants of the YB-1 cDNA (Fig. 4*A*). Clone 1b4 lacked 96 internal amino acid residues in the C terminus (from residue 158 to residue 245 of YB-1; ref. 22), while clone 28-1a lacked 108 amino acid residues (from residue 158 to 256). Clone 5b4 was identical to the *dbpA* gene (GenBank accession no. M24069). These findings revealed that OctT3 is a member of the Y-box family of transcription factors. A comparison of the relative intensities with which the *lacZ*-fusion proteins bound the BSO8 and wt-Oct probes (Fig. 4*B*) confirmed that the

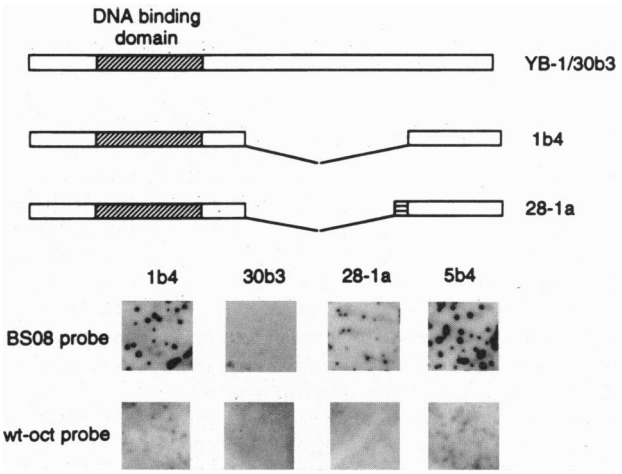


FIG. 4. Structure and binding specificities of the putative OctT3 cDNAs. (*A*) Four cDNA clones were isolated by screening a Jurkat cDNA library by the southwestern protocol with the BSO8 probe to detect site-specific DNA–protein interactions. Clones 30b3 and 5b4 were identical to YB-1 and dbpA genes, respectively. 1B4 and 28-1a are putative splice variants of YB-1. (*B*) Comparison of the relative intensities with which the *lacZ*-fusion polypeptides bound to the BSO8 (*Upper*) or wt-Oct (*Lower*) probes.

BSO8 probe is a preferred binding site over the wt-Oct probe. Because of the low affinities of the *lacZ*-fusion proteins for the wt-Oct probe, they may easily have gone undetected in previous screens with the Octamer probe. Thus, the isolation of an optimal binding site for OctT3 by the present approach alleviated this problem and permitted its cloning.

We next employed a subset of the library probes to compare differences in factor binding activities in quiescent and activated Jurkat cells. The segment 6 (S6) probe showed two additional bands, termed S6b, with the HA-TP2 extract (Fig. 5, lane 2) that were absent with extracts from quiescent Jurkat cells (lane 1). These bands were competitively blocked by an excess of total S6 material (lane 3) and also by the extracted S6a probe(s) (lane 5). However, the S6b probe competed with the S6b bands efficiently, but only partially competed with the S6a complex (lane 6), suggesting that S6a could contain more than one comigrating DNA–protein complex. An NF-AT binding site (10), however, did not compete with any of the specific bands in the S6 probe (lane 4), although when used as a probe in EMSA it showed specific shifts that were competitively blocked by an excess of unlabeled NF-AT duplex but not by the total S6 material (not shown). This indicated that the S6b binding activities in activated nuclear extracts were distinct from those of NF-AT.

Sequences from 30 clones contained in the S6 probe (Table 2) fell into three different consensus patterns: 13 conformed to the consensus pattern RTGGGBGGRY (R = G or A; B = G, T, or C; and Y = T or C) (S6-1), 8 to the consensus pattern YKRTAAMS (K = T or G, M = A or C, and S = G or C) (S6-2), and 5 to the consensus pattern WCCGGAAGY (W = A or T) (S6-3). Synthetic oligonucleotides containing the consensus core sites were used in the EMSA analysis. The S6-3 synthetic oligonucleotide duplex, but not S6-1 or S6-2, competed with the S6b bands (Fig. 5*B*, lane 4). Conversely, the shifts produced by the synthetic S6-3 probe were competitively blocked equally well by S6-3 and S6b (Fig. 5*B*, lanes 8 and 9, respectively) but not by the other synthetic probes, indicating that the S6b bands with activated extracts were due to the S6-3 consensus sequence. Comparison of the S6-3 consensus sequence with the data bases revealed striking homologies with binding sites for the Ets family of transcription factors (23, 24).

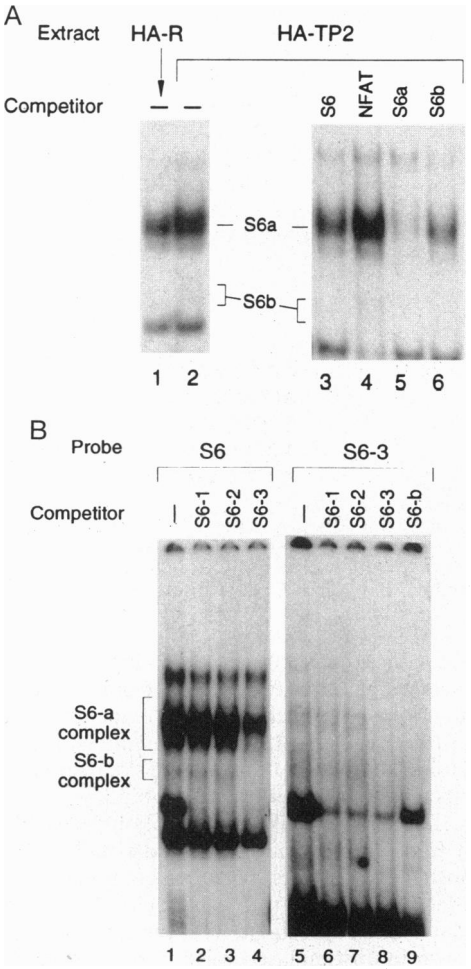


FIG. 5. Appearance of additional Ets-like binding activities in Jurkat cells treated with phorbol 12-myristate 13-acetate/phytohemagglutinin. (*A*) An EMSA was performed with the S6 probe by using nuclear extracts from either resting (HAR, lane 1) or activated Jurkat cells (HA-TP2, lane 2). The competitor oligonucleotides were total S6 material (lane 3), a synthetic NF-AT site (lane 4), the extracted S6a oligonucleotides (lane 5), or the extracted S6b oligonucleotides (lane 6). (*B*) Competition of the S6b binding activity by the synthetic S6-3 consensus probe (see Table 2). An EMSA experiment was done by using either the S6 probe (lanes 1–4) or the synthetic consensus S6-3 probe (lanes 5–9). The competitor oligonucleotides were S6-1 (lanes 2 and 6), S6-2 (lanes 3 and 7), S6-3 (lanes 4 and 8), or the extracted S6b oligonucleotides (lane 9).

## DISCUSSION

Several approaches for the selection of random oligonucleotides have been described that were aimed at determining the target recognition sites of DNA-binding proteins (25–27). These methods have involved the use of single target macromolecules or nucleic acids for selection of randomized oligonucleotides. The consensus DNA binding sites for several proteins have been determined by using these approaches (28). A somewhat different approach, termed CASTing (29), has been used to identify transcription factors that bind adjacent to and cooperate with a specific factor for which an antibody is available. The goal of the approach we present here is different from the above approaches in that we designed experiments to obtain the binding sites for multiple factors simultaneously without prior knowledge of the factors. Since several independent DNA–protein interactions occur simultaneously in this approach, we have termed it "Multiplex Selection Technique" (MuST). Success in selecting the highest affinity sites with the MuST approach is dependent upon the use of optimum conditions that define the ratio of specific to nonspecific sites selected for each interaction (25).

The molar concentrations of an asymmetric 10-mer binding site in our experiments is $5 \times 10^{-15}$, of a 6-mer site is $5 \times 10^{-12}$, and of a protein is estimated to be $3.5 \times 10^{-10}$ (for a protein of 40 kDa, which may be present at an abundance level of 1:10,000 with respect to the total mass of protein in the extract). Under these conditions, the concentration of probe and a single protein will be well below the average $K_d$ values (assumed to be $10^{-8}$ M). Therefore, the binding of any target is not driven to saturation, thus permitting discrimination of optimal and suboptimal sites. At the same time, because there are multiple protein targets in the same mixture, the total amount of oligonucleotide selected specifically by some protein will be relatively high compared with the background of nonspecifically retained probe. The ratios of binding sites selected by two proteins will be determined by $[(P_1/P_2)(K_1/K_2)]^n$, where $P_1$ and $P_2$ and $K_1$ and $K_2$ are their respective concentrations and binding constants, and $n$ is the number of cycles of selection. Because of the large number of binding proteins, the total ratio of specifically to nonspecifically retained oligonucleotides will be relatively high at each cycle, and a limited number of cycles of selection are needed to get a library in which most oligonucleotides represent specific binding sites.

By using relatively short random oligonucleotides, we decrease the chance of selecting oligonucleotides that contain separate binding sites for two different proteins. This avoids information or confusion about interaction between proteins separately bound to DNA. In turn, this simplifies the analysis of data and cataloging of DNA binding sites. The sequences of selected binding activities examined in this study showed a small number of extended binding sites rather than a wide spread of suboptimal sites (see Fig. 3*B* and Table 2). The use of native proteins in our approach and in CASTing (29) has advantages over exogenously synthesized proteins. Accessory protein factors may modify the stringency and selectivity of some DNA-binding proteins (30–32). Also, protein–protein interactions might create a protein–DNA interaction motif where neither partner may bind to DNA independently or might stabilize binding of a factor.

At a minimum, 8 of 20 clones picked at random from the library of binding sites resembled recognition sites for known transcription factors. Two sequences resembled binding sites for Pit-1 (17), a POU-homeodomain transcription factor that is preferentially expressed in the pituitary gland (33) that has recently been detected also in hematopoietic and lymphoid cells (34). A third sequence resembled the binding site for ANF, a negative regulator of the albumin enhancer in hepatocytes (16), whose presence in T cells has not been demonstrated. Therefore, we analyzed by EMSA these three sites and four other sites that did not match with sequences in the data bases. BSN17, one of the Pit-1-like sites, exhibited two complexes in Jurkat cells (Fig. 1) but a single complex in JY cells (not shown), revealing the presence in lymphoid cells of Pit-1-like binding activities. The ANF-like site showed a single complex. Further, two of the four unknown sites showed specific shifts. Taken together, these data demonstrate that at least 70% of the DNA probes contained in the binding site library represent demonstrable binding sites.

Accurate estimation of the total complexity of binding sites in the library was not possible. However, the low level of even imperfect repeats of binding site sequences in randomly sequenced clones suggested that hundreds of sites may be represented. Also, when multiple clones with the same binding site were sequenced, as in the case of the Ets-like sequences, the structure of individual clones were all different, indicating that clones were not selectively amplified from an extremely small pool by a "Monte Carlo" effect.

Table 2. Three distinct consensus patterns of oligonucleotide sequences in segment 6

| | | | |
|---|---|---|---|
| 6-2-2 | GGGG**ATGGGCgg**atcc | 6-1 | AATT**GTAA**GC |
| 6-3-1 | TAGT**GTGGGTgg**atcc | 6-2-1 | TTAG**GTAA**AT |
| 6-3-3 | ATTT**ATGGGGgg**atcc | 6-5-2 | CGCT**GTAA**TA |
| 6-5-1 | GTAG**GTGGGCgg**atcc | 6-6-2R | CC**G TAA**AGGG |
| 6-5-3R | CGGAA**GTGGCcga**tacc | 6-7-5R | GCCT**ATAACG** |
| 6-6-1 | GACAC**GTGATgga**tcc | 6-9-2 | GGG**ATAAACGC** |
| 6-6-3 | T**ATGGGGGA**A | 6-10-1R | CAT**CTAA**CC |
| 6-7-1 | **GTGGGTGG**GC | 6-10-2 | CATC**ATAA**CC |
| 6-9-1 | GATT**ATGGGTgg**atcc | Consensus | |
| 6-9-3 | G**GTTGGGGG**C | S6-2 | Yt/gRTAAa/cc/g |
| 6-11-2R | **GTGGGCGG**GG | | |
| 6-12-2R | **GTGGGGGC**TC | | |
| 6-12-3 | TGTC**GTGGGTgg**atcc | 6-3-2R | A**ACCGGAAGT** |
| Consensus | | 6-4-1R | CGT**ACCGGAAgg** |
| S6-1 | RTGGGBGGRY | 6-8 | **TCCGGAAGC** |
| | | 6-11-1 | **TCCGGAAGC** |
| | | 6-11-3 | **ACCGGAAGT** |
| | | Consensus | |
| | | S6-3 | a/tCCGGAAGY |

Core sequences are indicated in boldface type, and heel sequences are in lowercase letters; boldface lowercase letters belong to the heel sequence. R, purine; Y, pyrimidine; B, C, T, or G; K, T or G; M, A or C; S, C or G; W, A or T.

Nonspecific interactions might arise because of binding of single-stranded or double-stranded DNA, ends of DNA molecules, or hairpin loops by proteins present in the nuclear extracts. The large excess of poly(dI-dC) and heparin-agarose fractionation of nuclear extracts might have reduced these interactions to some extent. Certain intense bands were seen with one set of PCR heel sequences (3′-terminal sequence; see Table 2) but not with a second set, and vice versa (not shown). The PCR heel sequences may also contribute a part of some binding sites. In the present study one set of sequences, S6-1 (Table 2), almost certainly involved partial binding to the PCR heel sequences, since the core sequences were all adjacent to one of the heels and the synthetic core site did not exhibit EMSA shifts in absence of the heel sequence (data not shown).

As described above, we isolated a mixture of binding sites that strongly competed with the OctT3 EMSA band produced with a wt-Oct probe. These sequences were used to isolate cDNA clones of the Y-box family, a group of proteins previously cloned because of their interaction with the Y-box of the human major histocompatibility complex class II *HLA-DRA* chain promoter (22). The sequences we isolated were categorized into two consensus patterns, an A+T-rich set (BSO8) similar to the sequences in and flanking the Y-box, and a G+C-rich set (BSO3/9) homologous to sequences in and flanking the adjacent X-box of the *HLA-DRA* promoter (Fig. 3*B*) (22). The reason(s) for seeing multiple EMSA complexes with the BSO8 probe and only a single complex with the BSO3/9 probe are unclear. Recently, YB-1 was shown to bind to regions of the X-box and the Y-box (35). Curiously, a consensus octamer site is present adjacent to the Y-box motif in the *DRA* promoter.

Our data suggest that the OctT3 activity could be a member of the "cold shock" family of transcription factors. The successful cloning of the OctT3 binding activity by the alternate method employed here demonstrates the selectivity of MuST for naturally occurring sites. The finding that the BSO8 probe bound more strongly than the wt-Oct probe to the LacZ polypeptides reconfirms our predictions for the presence of optimal probes in the library.

Partitioning the library probes into gel segments facilitated detailed analysis by decreasing the complexity. With some exceptions, the pattern of bands with any segment probe was identical in resting and activated Jurkat nuclear extracts. However, segment 6 showed additional bands in activated extracts as compared with the resting extracts. These differences were consistently observed in duplicate experiments with extracts prepared at two different times and in heparin-agarose-fractionated as well as in crude nuclear extracts from activated Jurkat cells. EMSA experiments with synthetic S6 consensus sites showed that the binding site responsible for the S6b activity, in fact, resembles an Ets binding site (23).

A large family of Ets-like transcription factors have been described. Ets-1 and Ets-2 are expressed at high levels in T cells (36). Another set of Ets proteins comprises the ternary complex factors (TCFs) (37)—namely, Elk-1, Sap-1a, and Sap-2—that remain associated with the c-fos serum response element even in unstimulated cells; transcriptional activation by them is potentiated by phosphorylation in response to serum, v-ras, phorbol 12-myristate 13-acetate, and other agents. Ets-like factor Elf-1, a T-cell-specific Ets protein, has been demonstrated to be involved in the expression of a number of genes (38). Although inducible by mitogens in peripheral T cells, Elf-1 expression was found to be constitutive in Jurkat and other lymphoid cell lines (38). One component of the S6a band in our studies is a constitutively expressed factor whose relative mobility in EMSA gels is consistent with the molecular size of Elf-1. However, the S6b bands might represent yet unidentified members of the Ets family. Ets-like activities with comparable mobilities in EMSA gels to those of the S6b complexes have been observed (39, 40) that were not supershifted by antisera to either Ets-1, Elf-1, or GA-binding proteins, GABP$\alpha$ or GABP$\beta$. Our finding that the S6b activities are inducible in Jurkat cells suggests that additional members of the Ets family of transcription factors might exist in T cells.

Potentially, a cell-type-specific binding site library can be prepared by competition in the EMSA. Alternatively, consensus binding site sequences derived from a statistical analysis resulting from large-scale sequencing of binding sites from a single library may be used to determine cell-type-specificity of DNA recognition sites.

*A priori* one might have expected an overwhelmingly complex pattern in these experiments because of the large number of transcription factors, alternative splices, and complexes of several proteins. However, data presented here show that the actual pattern is workable. More importantly, since a large number of transcription factors can be analyzed *en masse* in a single study, the approach offers the investigator rapid insights into the distribution of factors in the cell type, and the identification of particular factors for further study, while

simultaneously generating optimal probes for cloning the factors.

We thank Drs. Ajay Bhargava, Patanjali Sankhavaram, and Yatindra Parashar for helpful comments, Tony Zhen Li for sequence analysis, and the Keck Biotechnology Resource Laboratory for synthetic oligonucleotides. Financial support from a grant (CA 42556) to S.M.W. is gratefully acknowledged.

1. Akam, M. (1987) *Development (Cambridge, U.K.)* **101,** 1–22.
2. Buckingham, M. E. (1994) *Curr. Opin. Genet. Dev.* **4,** 745–751.
3. Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N. & Weintraub, H. (1995) *Science* **268,** 836–844.
4. Desprez, P. Y., Hara, E., Bissell, M. J. & Campisi, J. (1995) *Mol. Cell. Biol.* **15,** 3398–3404.
5. Bhargava, A. K., Li, Z. & Weissman, S. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10260–10264.
6. Gruffat, H. & Sergeant, A. (1994) *Nucleic Acids Res.* **22,** 1172–1178.
7. Schreiber, E., Matthias, P., Muller, M. M. & Schaffner, W. (1989) *Nucleic Acids Res.* **17,** 6419.
8. Wiederrecht, G., Shuey, D. J., Kibbe, W. A. & Parker, C. S. (1987) *Cell* **48,** 507–515.
9. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2650–2654.
10. Flanagan, W. M., Corthesy, B., Bram, R. J. & Crabtree, G. R. (1991) *Nature (London)* **352,** 803–807.
11. Kingston, R. E., Schuetz, T. J. & Larin, Z. (1987) *Mol. Cell. Biol.* **7,** 1530–1534.
12. Dynan, W. S. & Tjian, R. (1985) *Nature (London)* **316,** 774–778.
13. Evans, T., Reitman, M. & Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5976–5980.
14. Lee, W., Mitchell, P. & Tjian, R. (1987) *Cell* **49,** 741–752.
15. Tsukiyama, T., Niwa, O. & Yokoro, K. (1989) *Mol. Cell. Biol.* **9,** 4670–4676.
16. Herbst, R. S., Boczko, E. M., Darnell, J. E., Jr., & Babiss, L. E. (1990) *Mol. Cell. Biol.* **10,** 3896–3905.
17. Elsholtz, H. P., Albert, V. R., Treacy, M. N. & Rosenfeld, M. G. (1990) *Genes Dev.* **4,** 43–51.
18. Yang, Y. S., Hanke, J. H., Carayannopoulus, L., Craft, C. M., Capra, J. D. & Tucker, P. W. (1993) *Mol. Cell. Biol.* **13,** 5593–5603.
19. Messier, H., Brickner, H., Gaikwad, J. & Fotedar, A. (1993) *Mol. Cell. Biol.* **13,** 5450–5460.
20. Ullman, K. S., Flanagan, W. M., Edwards, C. A. & Crabtree, G. R. (1991) *Science* **254,** 558–562.
21. Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H. & McKnight, S. L. (1988) *Genes Dev.* **2,** 801–806.
22. Didier, D. K., Schiffenbauer, J., Woulfe, S. L., Zacheis, M. & Schwartz, B. D. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7322–7326.
23. Faisst, S. & Meyer, S. (1992) *Nucleic Acids Res.* **20,** 3–26.
24. Wang, C.-Y., Petryniak, B., Ho, I.-C., Thompson, C. B. & Leiden, J. M. (1992) *J. Exp. Med.* **175,** 1391–1399.
25. Irvine, D., Tuerk, C. & Gold, L. (1991) *J. Mol. Biol.* **222,** 739–761.
26. Blackwell, T. K. & Weintraub, H. (1990) *Science* **250,** 1104–1110.
27. Thieson, H.-J. & Bach, C. (1990) *Nucleic Acids Res.* **18,** 3203–3209.
28. Wright, W. E. & Funk, W. D. (1993) *Trends Biochem. Sci.* **18,** 77–80.
29. Funk, W. D. & Wright, W. E. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9484–9488.
30. Herskowitz, I. (1989) *Nature (London)* **342,** 749–757.
31. Yamamoto, K. R., Pearce, D., Thomas, J. & Miner, J. N. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 1169–1192.
32. Chang, C.-P., Shen, W.-F., Rozenfeld, S., Lawrence, H. J., Largman, C. & Cleary, M. L. (1995) *Genes Dev.* **9,** 663–674.
33. Lew, A. M. & Elsholtz, H. P. (1991) *Nucleic Acids Res.* **19,** 6329.
34. Delhase, M., Vergani, P., Malur, A., Hooghe-Peters, E. L. & Hooghe, R. J. (1993) *Eur. J. Immunol.* **23,** 951–955.
35. MacDonald, G. H., Itoh-Lindstrom, Y. & Ting, J. P.-Y. (1995) *J. Biol. Chem.* **270,** 3527–3533.
36. Tymms, M. J. & Kola, I. (1994) *Mol. Reprod. Dev.* **39,** 208–214.
37. Price, M. A., Rogers, A. E. & Treisman, R. (1995) *EMBO J.* **14,** 2589–2601.
38. John, S., Reeves, R. B., Lin, J.-X., Child, R., Leiden, J. M., Thompson, C. B. & Leonard, W. J. (1995) *Mol. Cell. Biol.* **15,** 1786–1796.
39. Wurster, A. L., Siu, G., Leiden, J. M. & Hedrick, S. M. (1994) *Mol. Cell. Biol.* **14,** 6452–6463.
40. Gunther, C. V. & Graves, B. J. (1994) *Mol. Cell. Biol.* **14,** 7569–7580.