

Affinity selection of DNA-binding protein complexes using mRNA display

Seiji Tateyama, Kenichi Horisawa, Hideaki Takashima, Etsuko Miyamoto-Sato, Nobuhide Doi and Hiroshi Yanagawa*

Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, 3-14-1, Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Received October 31, 2005; Revised December 21, 2005; Accepted January 27, 2006

ABSTRACT

Comprehensive analysis of DNA–protein interactions is important for mapping transcriptional regulatory networks on a genome-wide level. Here we present a new application of mRNA display for *in vitro* selection of DNA-binding protein heterodimeric complexes. Under improved selection conditions using a TPA-responsive element (TRE) as a bait DNA, known interactors *c-fos* and *c-jun* were simultaneously enriched about 100-fold from a model library (a 1:1:20 000 mixture of *c-fos*, *c-jun* and *gst* genes) after one round of selection. Furthermore, almost all kinds of the AP-1 family genes including *c-jun*, *c-fos*, *junD*, *junB*, *atf2* and *b-atf* were successfully selected from an mRNA display library constructed from a mouse brain poly A⁺ RNA after six rounds of selection. These results indicate that the mRNA display selection system can identify a variety of DNA-binding protein complexes in a single experiment. Since almost all transcription factors form heterooligomeric complexes to bind with their target DNA, this method should be most useful to search for DNA-binding transcription factor complexes.

INTRODUCTION

The specific interactions between *cis*-regulatory DNA elements and transcription factors are critical components of transcriptional regulatory networks (1,2). The whole genome and complete cDNA sequences contain a large number of transcription factors and their binding DNA sequences, and thus comprehensive analysis of DNA-transcription factor interactions is expected to provide a deep understanding of the mechanisms of cell proliferation, developmental processes in tissue morphogenesis and disease. Currently, combined use

of chromatin immunoprecipitation (ChIP) assay with DNA-microarrays (ChIP-chip) (3–5) is the most widely used high-throughput method for discovering *cis*-regulatory DNA elements for a transcription factor. In contrast, development of high-throughput methods for discovering transcription factors for a *cis*-regulatory DNA element remains at an early stage. Although the yeast one-hybrid method (6,7) and phage display (8–10) are attractive candidates, these methods are not easily scalable because of the use of living cells. In addition, as over-expression of transcription factors often affects cellular metabolisms, such transcription factors are difficult to screen. In order to circumvent these difficulties, we focused on a totally *in vitro* mRNA display technology (11–17), *in vitro* virus (IVV) (11–14), for the discovery of DNA–protein interactions. In mRNA display, a library of genotype (mRNA)–phenotype (protein) linking molecules (IVV) is constructed in which mRNA is covalently bound to protein through puromycin during cell-free translation. After affinity selection via the protein moiety of the IVV, the mRNA moieties of the selected molecules are amplified by means of RT–PCR. Therefore, even very low-copy number proteins can be identified by iterative affinity selection from a library with high diversity and complexity, routinely in the range of 10¹³ members (11–17).

We previously demonstrated that the IVV selection system is effective for selection of protein–protein interactions in the case of c-Jun and c-Fos bait proteins (13). In this study, we show for the first time that the IVV selection system is also useful for selection of DNA–protein complex interactions (Figure 1). As a model bait DNA, we chose a TPA-responsive element (TRE; TGAC/GTCA), which is a common feature of promoter and enhancer sequences of many mammalian genes, such as collagenase I, SV40, interleukin 2, CD44 and TNF α (18). We show here that IVVs for c-Fos and c-Jun can form a heterodimer complex and interact with TRE in a sequence-specific manner, and that almost all kinds of the AP-1 family proteins can be enriched and selected with TRE-immobilized beads from an IVV library constructed from a mouse brain cDNA library.

*To whom correspondence should be addressed. Tel: +81 45 566 1775; Fax: +81 45 566 1440; Email: hyana@bio.keio.ac.jp

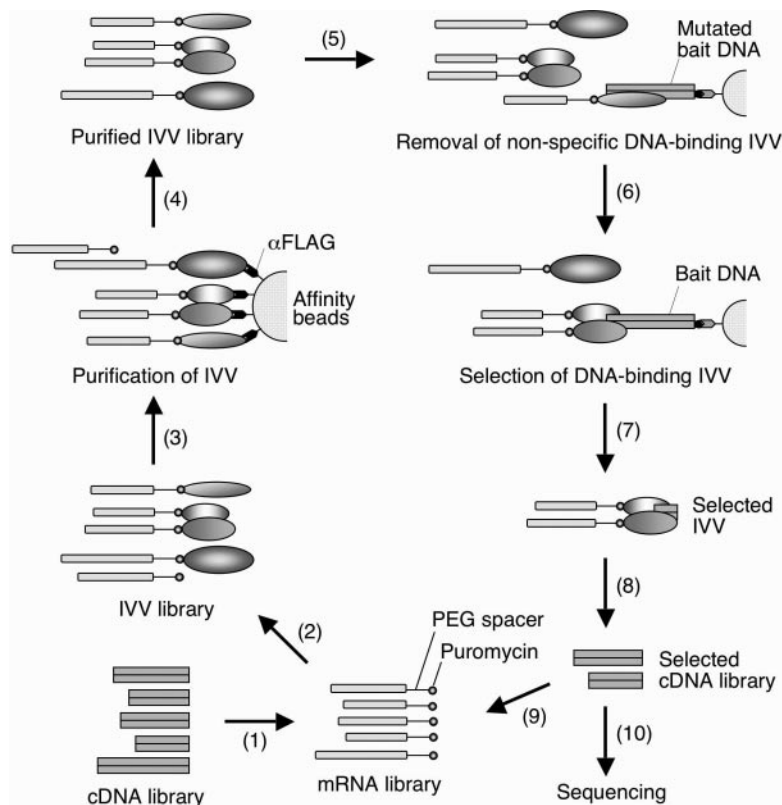


Figure 1. Schematic representation of the IVV selection procedure. (1) A cDNA library is transcribed and ligated with a Fluoro-PEG Puro spacer. (2) The IVV template RNA library is translated in a wheat germ cell-free translation system. (3) The constructed IVV library is purified with anti-FLAG M2 antibody-immobilized beads to remove untranslated mRNA and impurities contained in the wheat germ mixture. (4) The purified IVV are (5) subjected to affinity selection with mutated bait DNA-immobilized beads to eliminate non-specific binders to DNA. (6) Unbound IVV are then subjected to affinity selection with bait DNA-immobilized beads. (7) After washing, DNA-binding IVV are eluted by DNase I digestion. (8) The mRNA portions of the selected IVV are reverse-transcribed, PCR-amplified and (9) subjected to the next round of selection or (10) identified by cloning and sequencing.

MATERIALS AND METHODS

Preparation of bait DNA-immobilized affinity beads

Bait DNA was prepared by hybridization of two complementary chemically synthesized oligonucleotide DNAs (19–22) (Supplementary Table 1). Briefly, a mixture of equal amounts of oligonucleotides dissolved in DNA-binding buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 M NaCl and 0.1% Triton X-100] was mixed, heated at 95°C for 10 min, and cooled to room temperature for 2–3 h. Then, 150 pmol of the DNA mixture was added to 20 μ l of either Streptavidin- or NeutrAvidin-immobilized agarose beads (Pierce) and incubated at 4°C for 1 h. The beads were washed with DNA-binding buffer twice and equilibrated with selection buffer [50 mM Tris-HCl (pH 7.6), 50 mM NaCl, 1 mM EDTA, 5 mM MgCl₂, 5 mM DTT and 2% glycerol].

Preparation of positive and negative control IVV constructs

DNA templates for c-Fos (encoding c-Fos protein, amino acids 118–212), c-Jun (encoding c-Jun protein, amino acids 237–334), and glutathione *S*-transferase (GST) (encoding GST protein, amino acids 1–160) IVV formation were prepared by means of two PCR steps using the primers listed in Supplementary Table 1. The first PCR was performed using

either a mouse brain cDNA library (Takara Shuzo) or a pGEX-4T-3 GST fusion vector (Amersham) as templates with the primers 5'T7c-Fos and 3'c-FosFlag, 5'T7c-Jun and 3'c-Jun-Flag, and 5'T7GST and 3'GSTFlag, respectively. The second PCR was performed using the first PCR products as templates with the primers 5'FWT7 containing SP6 promoter, translational enhancer from tobacco mosaic virus (12–14), and T7 tag and 3'RV30 containing Flag tag and A tail. After purification with a QIAquick PCR purification kit (Qiagen), the DNA templates were transcribed with a RiboMAX large-scale RNA production system SP6 (Promega). The resulting mRNA was purified with an RNeasy RNA purification kit (Qiagen) and ligated to a Fluoro-PEG Puro spacer [p(dCp)₂-T(Fluor)p-PEGp-(dCp)₂-puromycin] (12–14) with a T4 RNA ligase (Takara Shuzo). The ligated mRNA was again purified with the RNeasy RNA purification kit.

Preparation of IVV template library

An IVV template library was prepared according to our previous protocols (12–14) with some modifications. Briefly, 1 μ g of mouse brain polyA⁺ mRNA (BD Biosciences Clontech) was reverse-transcribed using a SuperScript double-strand cDNA synthesis kit (Invitrogen) and 2 pmol of a 3' random primer (Supplementary Table 1) according to the manufacturer's instructions. The cDNA was ligated with an adaptor DNA

(Supplementary Table 1) using a Ligation High kit (Toyobo). The ligated cDNA was purified with the QIAquick PCR purification kit and PCR-amplified using 5'FW and 3'RV36 primers (Supplementary Table 1) and a TaKaRa Ex Taq hot start version (Takara Shuzo). The PCR product was purified with the QIAquick PCR purification kit, size-fractionated with a CHROMA SPIN-1000 (BD Biosciences Clontech), and transcribed in the same manner as described above. The resulting mRNA was purified with the RNeasy RNA purification kit, size-fractionated with the CHROMA SPIN-1000, and ligated to the Fluoro-PEG Puro spacer with the T4 RNA ligase. The ligated mRNA was again purified with the RNeasy RNA purification kit.

Affinity selection

IVV formation was performed as described previously (12–14) with some modifications. A 100 μ l aliquot of a PROTEIOS wheat germ cell-free protein synthesis mixture (Toyobo), containing 20 pmol of the ligated RNA and 80 U of a SuperaseIn RNase inhibitor (Ambion), was incubated at 24°C for 2 h. The constructed IVVs were added to 60 μ l of anti-FLAG M2 antibody-immobilized agarose beads (Sigma) equilibrated with 40 μ l of FLAG binding buffer [50 mM HEPES-NaOH (pH 8.3), 150 mM NaCl and 0.25% Triton X-100] and mixed on a rotator at 4°C for 1 h. The beads were washed with 400 μ l of FLAG binding buffer four times and treated with 50 μ l of FLAG elution buffer [50 mM Tris-HCl (pH 7.6), 50 mM NaCl, 1 mM EDTA, 5 mM MgCl₂, 2% glycerol, 5 mM DTT, 1 μ g of poly-d(I-C) and 1 μ g/ μ l of FLAG M2 peptide (Sigma)] at 4°C for 1 h. In the case of selection from a mouse brain cDNA library, the purified IVVs were pre-selected with mutated TRE bait DNA-immobilized agarose beads (Supplementary Table 1) (21). After incubation at 4°C for 1 h, the flow-through fraction containing unbound IVVs was subjected to affinity selection.

The resulting IVVs were added to the bait DNA-immobilized agarose beads and mixed on a rotator at 4°C for 2 h. Streptavidin and NeutrAvidin agarose beads were alternately used to avoid enrichment of Streptavidin- or NeutrAvidin-specifically bound molecules. The beads were washed with 50 μ l of the selection buffer (described above) seven times followed by 50 μ l of DNase buffer [40 mM Tris-HCl (pH 8.0), 10 mM NaCl, 10 mM CaCl₂ and 6 mM MgCl₂] once. Then, 50 μ l of DNase buffer containing 3 U of DNase I (Promega) was added and the mixture was incubated at room temperature for 10 min. The reaction was terminated by addition of 8 μ l of 0.25 M EGTA. The resulting supernatant was used as a template for RT-PCR. The RT-PCRs were performed with a OneStep RT-PCR kit (Qiagen) using 5'FW and 3'RV30 primers (Supplementary Table 1). The resulting RT-PCR product was used for the next round of selection as described above or analyzed by quantitative real-time PCR.

Quantitative real-time PCR analysis

Real-time PCR was performed with a LightCycler FastStrand DNA master SYBR green I kit (Roche) using gene-specific primers (Supplementary Table 2) (13,18,22–28). The standard DNA was prepared by either PCR-amplification from the cloned plasmids using M13-FW and M13-RV primers

(Supplementary Table 1) or RT-PCR from the mouse brain polyA⁺ mRNA using gene-specific primers (Supplementary Table 2).

Cloning and sequencing

After selection, RT-PCR products were cloned using a Qiagen PCR cloning kit (Qiagen) and sequenced with an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The sequences were clustered using the CLUSTALW program and subjected to nucleotide–nucleotide BLAST (BLASTN) search to identify the protein represented by each cluster (13).

Pull-down assay

The DNA templates were prepared by PCR-amplification from the cloned plasmids using 5'FW and 3'RVXho27 primers (Supplementary Table 1). For c-Jun, c-Fos, JunD, JunB and B-ATF, the shortest clone was chosen in each case. The PCR products were transcribed, translated and affinity-purified as described in 'Affinity selection' except that DNase treatment was omitted. The bound proteins were eluted with SDS-PAGE sample buffer [0.1 M Tris-HCl (pH 6.8), 4% SDS, 0.2% bromophenol blue and 20% glycerol] at 90°C for 5 min and used for western blot analysis, which was performed with an ECF western blotting kit (Amersham Biosciences) and mouse anti-T7-tag monoclonal antibody (Novagen).

Electrophoretic mobility shift assay (EMSA)

EMSA was performed according to previous reports (20–22) with minor modifications. Cy5-labeled double-stranded probe DNA, non-specific double-stranded competitor DNA (21), and non-labeled double-stranded probe DNA (Supplementary Table 1) were prepared by hybridization as described above. Proteins produced from the DNA templates of selected clones were partially purified from the wheat germ cell-free protein synthesis mixture by using anti-FLAG M2 antibody-immobilized agarose beads (Sigma). The purity and concentration of the proteins were monitored on 10–20% SDS-PAGE with the ECF western blotting kit and the mouse anti-T7-tag monoclonal antibody. The assay mixture contained, in a total volume of 9 μ l, 50 mM Tris-HCl (pH 7.6), 50 mM NaCl, 1 mM EDTA, 5 mM MgCl₂, 5 mM DTT, 5% glycerol, an appropriate amount of proteins (~20 ng), and 0.5 μ g of poly-d(I-C). The mixture was incubated at room temperature for 5 min. Then, 100 fmol of Cy5-labeled probe DNA was added and incubation was continued for 20 min. In the case of competition assay, either non-specific competitor DNA or non-labeled probe DNA was added to the mixture. The samples were electrophoresed on 4–20% gradient polyacrylamide gel (Daiichi Pure Chemicals, Japan) with 0.25 \times TBE buffer. Gels were run at 4°C for 2 h at 150 V and analyzed with a Molecular Imager FX (Bio-Rad Laboratories).

RESULTS AND DISCUSSION

Establishment of the improved IVV selection system

To determine whether the IVV selection system is effective for discovery of DNA-binding proteins (Figure 1), we first performed a model experiment using a bait DNA containing TRE and c-Fos and c-Jun proteins as prey. To obtain model

libraries, DNA templates for *c-fos*, *c-jun* and *gst* were mixed in ratios of 1:1:2 and 1:1:200. The model libraries were transcribed, ligated to the Fluoro-PEG Puro spacer, and translated in a wheat germ cell-free translation system. The resulting IVV were then incubated with the bait DNA-immobilized beads, washed and eluted by DNase I treatment. The eluates were amplified by RT-PCR and the products were analyzed by quantitative real-time PCR. As a result, both *c-fos* and *c-jun* were simultaneously enriched about 100-fold after one round of selection (Figure 2A, lanes 2 and 7), while no enrichment was observed in the absence of the bait DNA (Figure 2A, lanes 3 and 8), in the presence of 400 μ M free puromycin as an inhibitor of the formation of IVV (Figure 2A, lanes 4 and 9), and in the presence of a mutated TRE (CCGAATT) bait DNA (Figure 2A, lanes 5 and 10). When *c-fos* and *c-jun* were independently added to the model library, no enrichment and slight enrichment (about 10-fold) were observed, respectively (Figure 2B, lanes 5 and 8). These results indicate that the peptide portions of c-Fos and c-Jun IVVs can form a heterodimer complex and bind to the bait DNA in a sequence-specific manner. This observation is consistent with previous reports that c-Fos cannot bind to DNA and that a c-Fos/c-Jun heterodimer, rather than c-Jun/c-Jun homodimer, preferentially binds to DNA (18,19,21,22).

Furthermore, we performed an iterative selection using a 1:1:20 000 model library in the presence and absence of the bait DNA. After four rounds of selection in the presence of the bait DNA, *c-fos* and *c-jun* were enriched to a detectable level and *gst* was decreased (Figure 2C, lanes 2, 4, 6 and 8). The final ratio of *c-fos*, *c-jun* and *gst* was about 60:60:1, indicating that *c-fos* and *c-jun* had been enriched about 1.2×10^4 -fold versus *gst* (Figure 2D). On the other hand, no significant enrichment of *c-fos* and *c-jun* was observed in each round of selection in the absence of the bait DNA (Figure 2C, lanes 3, 5, 7 and 9). These results indicate that even low-copy number genes of DNA-binding proteins can be detected by repeating the selection round in the IVV selection system.

Selection of DNA-binding proteins from a cDNA library

Next we applied the established IVV selection system to the selection of TRE-binding proteins from a mouse brain cDNA library. Pre-selection of the cDNA-derived IVV library was performed using anti-FLAG M2 antibody-immobilized beads to remove untranslated mRNA and impurities contained in the wheat germ cell-free translation system mixture [Figure 1, step (3)] and using the mutated TRE bait DNA-immobilized beads to eliminate non-specific binders to DNA [Figure 1, step (5)]. Then, two different kinds of selection, i.e. selection in the presence of bait DNA [bait (+) selection], and selection in the absence of bait DNA [bait (-) selection], were performed to distinguish non-specific binders to the agarose beads (13,23). Selection was continued until almost all of the selected clones were found to be positives or positive candidates. After six rounds of the selection, the resulting libraries, 6th bait (+) and 6th bait (-), were cloned and 477 and 98 randomly chosen clones were sequenced, respectively. Clones found in both the bait (+) and the bait (-) libraries, or that contained stop codons, were removed. The remaining 451 clones were subjected to nucleotide-nucleotide BLAST

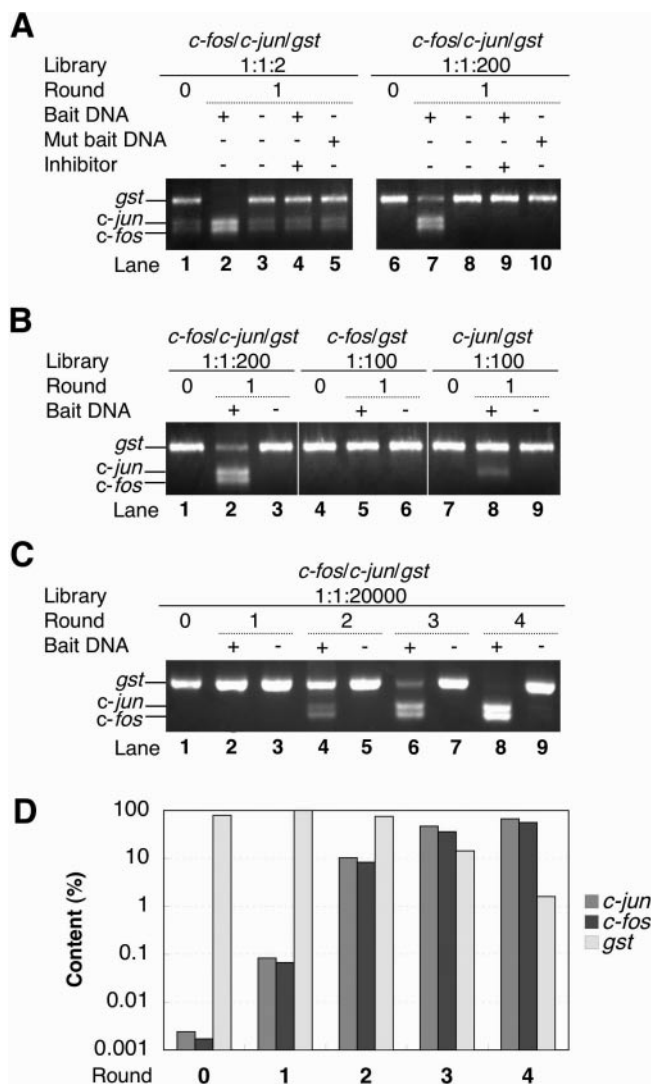


Figure 2. Affinity enrichment of DNA-binding protein complexes. (A) PCR profiles of the 1:1:2 and 1:1:200 model libraries are shown: initial library (lanes 1 and 6), libraries after one round of selection of the IVV library in the presence (lanes 2 and 7) and absence (lanes 3 and 8) of the bait DNA, libraries after one round of selection in the presence of 400 μ M free puromycin and the bait DNA (lanes 4 and 9) and libraries after one round of selection of the IVV library in the presence of the mutated TRE bait DNA (lanes 5 and 10). (B) PCR profiles of the *c-fos*, *c-jun* and *gst* (1:1:200), *c-fos* and *gst* (1:100), and *c-jun* and *gst* (1:100) model libraries are shown: initial library (lanes 1, 4 and 7), libraries after one round of selection in the presence (lanes 2, 5 and 8) and absence (lanes 3, 6 and 9) of the bait DNA. (C) PCR profile of each round of the 1:1:20 000 model library in the iterative selection. Initial library (lane 1) and libraries after first (lanes 2 and 3), second (lanes 4 and 5), third (lanes 6 and 7) and fourth (lanes 8 and 9) rounds of selection in the presence (+; lanes 2, 4, 6 and 8) and absence (-; lanes 3, 5, 7 and 9) of the bait DNA. (D) Contents of the *c-fos*, *c-jun* and *gst* genes in libraries at each round of the iterative selection. DNA (10^{10} molecules, 4–5 ng) was subjected to quantitative real-time PCR with specific primers (Supplementary Table 2).

(BLASTN) search to identify the encoded proteins. Of the 451 clones, 39 corresponded to either 3'-untranslated regions (3'-UTR) in mRNA sequences or incorrect reading frames. The remaining 412 clones were then clustered into six sequence groups and four individual sequences (Table 1).

The six sequence groups were known TRE-binding proteins, c-Jun, c-Fos, JunD, JunB, ATF2 and B-ATF. Each

Table 1. Characterization of the selected clones

Gene symbol	<i>c-jun</i>	<i>c-fos</i>	<i>junD</i>	<i>junB</i>	<i>atf2</i>	<i>b-atf</i>	hypothetical protein 1	<i>eef1a1</i>	<i>sgt1</i>	6330407J23Rik
Accession no.	NM_010591	NM_010234	NM_010592	NM_008416	NM_009715	NM_016767	CAA01058705	NM_010106	NM_026474	NM_026138
Number of clones	173	113	53	27	23	13	6	2	1	1
Types of selected protein fragments	10	4	3	4	1	2	1	1	1	1
Locus on mRNA	1556–1903	488–775	820–1131	971–1348	1265–1513	270–566	1728–1790 ^a	1408–1477	783–845	1864–1968
sequence (base)	1682–1858	551–763	892–1086	1115–1315	1265–1513	270–488	1728–1790 ^a	1408–1477	783–845	1864–1968
Locus on the shortest mRNA sequence (base)	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Leucine heptad repeats	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
TRE-binding motifs	+++	-/+++ ^c	+++	+++	+/+++ ^c	-/+++ ^c	+	+	+	-
Pull-down assay ^b	Yes	No/yes ^c	Yes	Yes	Yes	No/yes ^c	No	No	No	No
EMSA	Yes	No/yes ^c	Yes	Yes	Yes	No/yes ^c	No	No	No	No

^aLocus on *Mus musculus* whole genome shotgun assembly contig 47331.

^bThe relative binding activities are designated by plus and minus signs and represent the percentages of total protein that bound: +++, >10%; ++, 1–10%; +, 0.1–1%; -, not detected.

^cIn the presence of *c-jun*.

cluster except for ATF2 consists of plural protein sequences (Table 1). All of the six selected protein sequences contained the conserved basic region for TRE-interaction and the heptad repeat of leucine residues required for dimerization. To determine whether the selected protein fragments indeed specifically bind with the bait DNA, we performed pull-down assay and EMSA. As expected, *c-Jun*, *JunD* and *JunB* showed interaction with the bait DNA. ATF2 showed faint but significant interactions in pull-down assay and EMSA (Figure 3). The relative binding activity was enhanced in the presence of *c-Jun*. *c-Fos* and B-ATF themselves did not bind with the bait DNA, but did bind in the presence of *c-Jun* (Figure 3). In our experiments using EMSA, we found that binding affinities of *c-Fos/c-Jun/TRE*, *c-Fos/JunB/TRE*, *c-Fos/JunD/TRE* and *c-Jun/B-ATF/TRE* complexes ranged from 12.5 to 50 nM, whereas that of *c-Jun/ATF2/TRE* complex was about 200 nM. This result is not in conflict with previous reports on the selection of protein–protein interactions by mRNA display: the binding affinities of most selected proteins are <1 μM (16,17).

The four individual sequences were *Eef1a1*, *SGT1*, 6330407J23Rik and a hypothetical protein (named Hypothetical protein 1). Hypothetical protein 1 and 6330407J23Rik have not yet been registered as an open reading frame (ORF) in the public databases, though they are present in the mouse genome sequence and mRNA sequence, respectively. The other two are known proteins, in-frame, and within the native ORF, but have not been reported to bind with TRE, *c-Fos* or *c-Jun*. In DNA-binding experiments with these proteins, faint but possibly unspecific interactions were observed in pull-down assays, but no interaction was observed in EMSA (Figure 3). These interactions were not enhanced in the presence of other selected proteins. Search of the PSORTII program and other public databases revealed that *Eef1a1* and *SGT1* proteins are not located in the nucleus, implying that these proteins may not act as transcription factors. Although these four clones may be false-positives, further studies would be required to clarify whether Hypothetical protein 1 and 6330407J23Rik exist in cells and have biological significance, and whether these four proteins have a biologically relevant interaction with TRE *in vivo*.

Confirmation of the enrichment of other known candidates

On the basis that IVVs specifically interacting with TRE are enriched in the selection (13,23), we further performed quantitative real-time PCR analysis (23) to determine whether not only *c-jun*, *c-fos*, *junD*, *junB*, *atf2* and *b-atf*, but also other known positive candidates (18,24–28), *fosB*, *fra1*, *fra2*, *atf3* and *atf7*, were enriched in the 6th bait (+) library as compared with the initial library. As shown in Figure 4, significant enrichments were confirmed for *fosB*, *fra1*, *fra2* and *atf7* as well as all the selected clones listed in Table 1. None of the candidates was enriched in any round of the bait (–) libraries. In the case of *atf3*, significant enrichment could not be detected. The enrichment efficiencies of *fosB*, *fra1*, *fra2* and *atf7* were lower than those of *c-jun*, *c-fos*, *junD*, *junB*, *atf2* and *b-atf*, but higher than those of *hypothetical protein 1*, *eef1a1*, *sgt1* and 6330407J23Rik (Figure 4). The rates of successfully cloned known positives and successfully enriched known

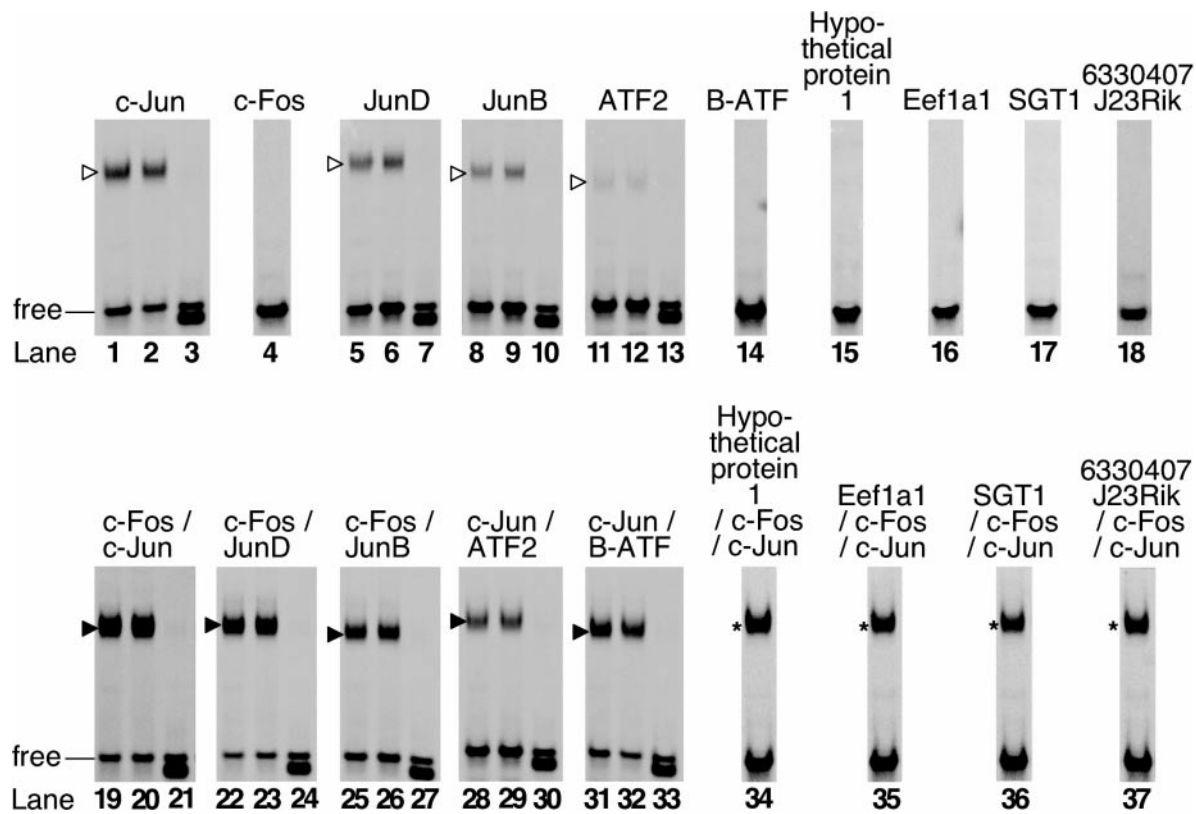


Figure 3. Interaction between the selected clones and TRE. Approximately 20 ng of the purified proteins were mixed with 100 fmol of the Cy5-labeled probe DNA and incubated at room temperature for 20 min (lanes 1, 4, 5, 8, 11, 14, 15, 16, 17, 18, 19, 22, 25, 28, 31, 34, 35, 36 and 37). Either 5 pmol (lanes 2, 6, 9, 12, 20, 23, 26, 29 and 32) of non-specific competitor DNA or 5 pmol (lanes 3, 7, 10, 13, 21, 24, 27, 30 and 33) of non-labeled probe DNA was added. Open arrowheads, homodimeric complexes; solid arrowheads, heterodimeric complexes; asterisks, a heterodimeric complex of c-Fos and c-Jun. The upper band of free probes (seen in all lanes) is a dsDNA consisting of two complementary oligonucleotide DNAs both labeled with Cy5 at their 5' ends, and the lower one (seen in lanes 3, 7, 10, 13, 21, 24, 27, 30 and 33) is an unexpectedly formed hybrid of non-labeled and labeled oligonucleotide DNAs.

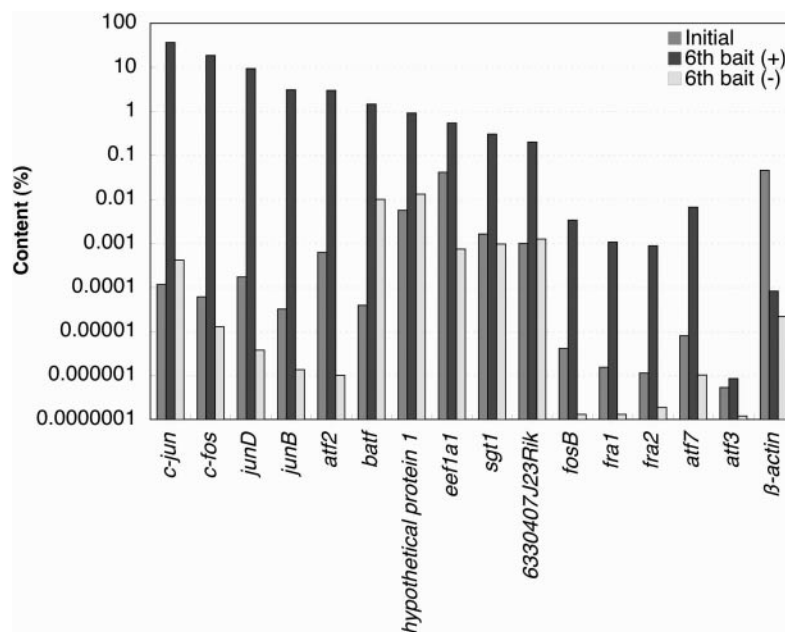


Figure 4. Quantitative real-time PCR analysis of the selected clones in the initial, 6th bait (+), and 6th bait (-) libraries. Contents of DNAs of selected clones and β -actin as a negative control in 5 ng (about 10^{10} molecules) of the each library were analyzed with specific primers (Supplementary Table 2) using a LightCycler. The enrichment rates of *c-jun*, *c-fos*, *junD*, *junB*, *atf2*, *b-atf*, *hypothetical protein 1*, *eef1a1*, *sgt1*, *6330407J23Rik*, *fosB*, *fra1*, *fra2* and *atf7* were 3.2×10^5 -, 3.0×10^5 -, 5.4×10^4 -, 9.6×10^4 -, 4.8×10^3 -, 3.7×10^4 -, 1.3×10^2 -, 13-, 1.6×10^2 -, 1.7×10^2 -, 9.2×10^2 -, 8.0×10^2 -, 8.7×10^2 -, and 8.3×10^2 -fold, respectively.

positives are 6/11 and 10/11, respectively. Although almost all these AP-1 family proteins (ten kinds) were significantly enriched, the question arises, why were c-Jun, c-Fos, JunD, JunB, ATF2 and B-ATF successfully cloned, while the others, so-called false-negatives, were not? As shown in Figure 4, the contents of *c-jun*, *c-fos*, *junD*, *junB*, *atf2*, *b-atf*, *fosB*, *fra1*, *fra2* and *atf7* in the 6th bait (+) library were about 37, 19, 9.4, 3.1, 3.0, 1.5, 0.0034, 0.0011, 0.00088 and 0.0066%, respectively. Because 412 clones were cloned and analyzed in this study, only genes whose content is more than 0.25% would be cloned theoretically. Indeed, the contents of all of the selected clones quantified by real-time PCR were at least 0.25%. Excess selection rounds do not result in enrichment of false-negatives, because the pool is dominated by the positives (such as *c-fos* and *c-jun*). If the affinities of false-negatives are much the same as those of the positives or lower than those of the positives, the false-negatives cannot exclude the positives from the saturated library. Thus, enrichment of such false-negative clones would be detected by the combination of the IVV system with a DNA microarray (29), as a sensitive and high-throughput alternative to the process of cloning and sequencing. Another possible reason for the occurrence of false-negatives may be their low contents (*fosB*, *fra1*, *fra2* and *atf7*, less than 0.00001%; *atf3*, less than 0.000001%), or low contents of in-frame genes (*fosB*, *fra1*, *fra2* and *atf7*, less than 0.0000017%, *atf3*, less than 0.00000017%, theoretically), in the initial library; in this case, the use of a normalized library constructed from tissues (30,31) or a full-length cDNA clone library may reduce the number of false-negatives.

In view of the enrichment factor of 100 for *c-jun* and *c-fos* found in the selection of model libraries (Figure 2), the question arises, why do those two proteins show only rather low enrichment factors of 3.2×10^5 and 3.0×10^5 , respectively, after a total of six rounds (Figure 4)? When the enrichment factors for *c-fos* and *c-jun* in the selection of the cDNA library were monitored in each round of selection, the factor of 50–170 in the 2nd round was comparable with that in the model selection, while the factors of 2–15 in the 1st and 3–6th rounds were relatively low. One possible reason for the lower enrichment efficiency in the 1st round of selection is that the initial library contains not only in-frame gene fragments, but also out-of-frame gene fragments from the same region as the in-frame gene fragments, because the IVV library was constructed by random-priming PCR of mRNA. Probably only in-frame gene fragments are enriched and out-of-frame gene fragments are removed in the 1st round of selection, and the selected in-frame gene fragments are then enriched in the latter rounds of selection. As the selection rounds proceed, DNA-binders other than *c-fos* and *c-jun* are also enriched in the pool and then compete with each other. In contrast, such competitors were absent in the model selection (only the *gst* gene was present).

Advantages of IVV selection system

The use of our IVV selection system for selection of DNA-protein interactions as described here has several advantages over previous techniques, such as one-hybrid and phage display systems, mainly due to the greater flexibility of the *in vitro* selection conditions and the high diversity and

complexity of the IVV library which contains about 10^{13} independent molecules. First, the IVV selection system is available for selection of DNA-binding protein heterodimeric complexes. Since almost all transcription factors form heterooligomeric complexes to bind with their target DNA, our results indicate that the IVV selection system would be more useful than previous techniques to search for DNA-binding transcription factor complexes. Phage display systems were recently utilized for selection of DNA-binding proteins (8–10), but no DNA-binding protein that forms a heterooligomeric complex was obtained. The yeast one-hybrid method (6,7) is not necessarily suitable for the detection of transcription factors that can form heterodimers and then bind with their target DNAs, but which cannot bind to the DNAs as monomeric proteins or homodimeric protein complexes. For example, c-Fos and B-ATF themselves do not show TRE-binding ability, and would be identified as negatives in the yeast one-hybrid system.

Second, the IVV selection system, a totally *in vitro* technique, is available for enrichment of many kinds of binders, such as c-Jun, c-Fos, JunD, JunB, ATF2, B-ATF, FosB, Fra1, Fra2 and ATF7 in a single experiment, probably because the diversity and complexity of the IVV library is not limited by the use of living cells. In particular, over-expression of transcription factors that are usually expressed at a low level is often toxic to the host cells in phage display and yeast one-hybrid systems.

Furthermore, the IVV selection system affords a remarkably low false-positive rate. Even if all four novel proteins are false-positives, the false-positive rate is only 2% (10 clones per 412 clones; Table 1). Purification using the anti-FLAG M2 antibody-immobilized beads and especially pre-selection of the IVV library with the mutated TRE bait DNA-immobilized beads were very effective to reduce both false-positives and false-negatives: when these processes were omitted, the number of false-positive clones reached ~50% and only *c-jun*, *c-fos* and *junD* were cloned (data not shown).

In summary, we have demonstrated the utility of the IVV system for selection, analysis and mapping of DNA-transcription factor interactions. The greater flexibility of the selection conditions of the IVV system, and the greater diversity and complexity of the IVV library allow easier selection of a variety of protein complexes with low rates of false-positives, as compared with current techniques such as phage display and the yeast one-hybrid method. Therefore, our system should contribute to the large-scale analysis of DNA-transcription factor interactions for mapping of transcriptional regulatory networks on a genome-wide level.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Ms A. Utsumi for experimental support. This work was supported in part by a Grant-in-Aid for Scientific Research, and a Special Coordination Fund from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Funding to pay the Open Access publication charges for this article was provided by Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

Conflict of interest statement. None declared.

REFERENCES

- Pollack, J.R. and Iyer, V.R. (2002) Characterizing the physical genome. *Nature Genet.*, **32**, 515–521.
- Yu, H., Luscombe, N.M., Quian, J. and Gerstein, M. (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.*, **19**, 422–427.
- Barski, A. and Frenkel, B. (2004) CHIP display: novel method for identification of genomic targets of transcription factors. *Nucleic Acids Res.*, **32**, e104.
- Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K. and Uhlmann, F. (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, **430**, 573–578.
- Kim, J., Bhinge, A.A., Morgan, X.C. and Iyer, V.R. (2005) Mapping DNA–protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nature Method.*, **2**, 47–53.
- Luo, Y., Vijaychander, S., Stile, J. and Zhu, L. (1996) Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one-two-hybrid systems. *Biotechniques*, **20**, 564–568.
- Lemerrier, C., To, R.Q., Swanson, B.J., Lynos, G.E. and Konieczny, S.F. (1997) Mist1: a novel basic helix-loop-helix transcription factor exhibits a developmentally regulated expression partner. *Dev. Biol.*, **182**, 101–113.
- Hagiwara, H., Kunihiro, S., Nakajima, K., Sano, M., Masaki, H., Yamamoto, M., Pak, J.W., Zhang, Y., Takase, K., Kuwabara, I. *et al.* (2002) Affinity selection of DNA-binding proteins from yeast genomic DNA libraries by improved λ phage display vector. *J. Biochem.*, **132**, 975–982.
- Cicchini, C., Ansuini, H., Amicone, L., Alonzi, T., Nicosia, A., Cortese, R., Tripodi, M. and Luzzago, A. (2002) Searching for DNA–protein interactions by lambda phage display. *J. Mol. Biol.*, **322**, 697–706.
- Rebar, E.J. and Pabo, C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specifications. *Science*, **263**, 671–673.
- Nemoto, N., Miyamoto-Sato, E., Hushimi, Y. and Yanagawa, H. (1997) *In vitro* virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. *FEBS. Lett.*, **414**, 405–408.
- Miyamoto-Sato, E., Takashima, H., Fuse, S., Sue, K., Ishizaka, M., Tateyama, S., Horisawa, K., Sawasaki, T., Endo, Y. and Yanagawa, H. (2003) Highly stable and efficient mRNA templates for mRNA–protein fusions and C-terminally labeled proteins. *Nucleic Acids Res.*, **31**, e78.
- Horisawa, K., Tateyama, S., Ishizaka, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Doi, N. and Yanagawa, H. (2004) *In vitro* selection of Jun-associated proteins using mRNA display. *Nucleic Acids Res.*, **32**, e169.
- Miyamoto-Sato, E., Ishizaka, M., Horisawa, K., Tateyama, S., Takashima, H., Fuse, S., Sue, K., Hirai, N., Masuoka, K. and Yanagawa, H. (2005) Cell-free co-translation and selection using *in vitro* virus for high-throughput analysis of protein–protein interactions and complexes. *Genome Res.*, **15**, 710–717.
- Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl Acad. Sci. USA*, **94**, 12297–12302.
- Hammond, P.W., Alpin, J., Rise, C.E., Wright, M. and Kreider, B.L. (2001) *In vitro* selection and characterization of Bcl-X_L-binding proteins from a mix of tissue-specific mRNA display libraries. *J. Biol. Chem.*, **276**, 20898–20906.
- Shen, X., Valencia, C.A., Szostak, J.W., Dong, B. and Liu, R. (2005) Scanning the human proteome for calmodulin-binding proteins. *Proc. Natl Acad. Sci. USA*, **102**, 5969–5974.
- Chinenov, Y. and Kerppola, T.K. (2001) Close encounters of many kinds: Fos–Jun interactions that mediate transcription regulatory specificity. *Oncogene*, **20**, 2438–2452.
- Abate, C., Luk, D., Gentz, R., Rauscher, F.J.III. and Curran, T. (1990) Expression and purification of the leucine zipper and DNA-binding domains of Fos and Jun: both Fos and Jun contact DNA directly. *Proc. Natl Acad. Sci. USA*, **87**, 1032–1036.
- Doi, N., Takashima, H., Kinjo, M., Sakata, K., Kawahashi, Y., Oishi, Y., Oyama, R., Miyamoto-Sato, E., Sawasaki, T., Endo, Y. *et al.* (2002) Novel fluorescence labeling and high-throughput assay technologies for *in vitro* analysis of protein interactions. *Genome Res.*, **12**, 487–492.
- John, M., Leppik, R., Busch, S.J., Granger-Schnarr, M. and Schnarr, N. (1996) DNA binding of Jun and Fos bZIP domains: homodimers and heterodimers induce a DNA conformational change in solution. *Nucleic Acids Res.*, **24**, 4487–4494.
- Ramidez-Carrozzi, V. and Kerppola, T. (2003) Asymmetric recognition of nonconsensus AP-1 sites by Fos-Jun and Jun-Jun influences transcriptional cooperativity with NFAT1. *Mol. Cell Biol.*, **23**, 1737–1749.
- Horisawa, K., Doi, N., Takashima, H. and Yanagawa, H. (2005) Application of quantitative real-time PCR for monitoring the process of enrichment of clones on *in vitro* protein selection. *J. Biochem.*, **137**, 121–124.
- Hai, T. and Curran, T. (1991) Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc. Natl Acad. Sci. USA*, **88**, 3720–3724.
- Echlin, D.R., Tae, H.J., Mitin, N. and Taparowsky, E.J. (2000) B-ATF functions as a negative regulator of AP-1 mediated transcription and blocks cellular transformation by Ras and Fos. *Oncogene*, **19**, 1752–1763.
- Suzuki, T., Okuno, H., Yoshida, T., Endo, T., Nishina, H. and Iba, H. (1991) Difference in transcriptional regulatory function between c-Fos and Fra-2. *Nucleic Acids Res.*, **19**, 5537–5542.
- Zerial, M., Toschi, L., Ryseck, R.P., Schuermann, M., Muller, R. and Bravo, R. (1989) The product of a novel growth factor activated gene, fos B, interacts with JUN proteins enhancing their DNA binding activity. *EMBO J.*, **8**, 805–813.
- Peters, C.S., Liang, X., Li, S., Kannan, S., Peng, Y., Taub, R. and Diamond, R.H. (2001) ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. *J. Biol. Chem.*, **276**, 13718–13726.
- Mockler, T.C. and Ecker, J.R. (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics*, **85**, 1–15.
- Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.*, **10**, 1617–1630.
- Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A. *et al.* (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res.*, **32**, e37.