Cell-free cotranslation and selection using in vitro virus for high-throughput analysis of protein–protein interactions and complexes

Etsuko Miyamoto-Sato, Masamichi Ishizaka, Kenichi Horisawa, Seiji Tateyama, Hideaki Takashima, Shinichiro Fuse, Kaori Sue, Naoya Hirai, Kazuyo Masuoka, and Hiroshi Yanagawa[1]

Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

We have developed a simple and totally in vitro selection procedure based on cell-free cotranslation using a highly stable and efficient in vitro virus (IVV). Cell-free cotranslation of tagged bait and prey proteins is advantageous for the formation of protein complexes and allows high-throughput analysis of protein–protein interactions (PPI) as a result of providing in vitro instead of in vivo preparation of bait proteins. The use of plural selection rounds and a two-step purification of the IVV selection, followed by in vitro post-selection, is advantageous for decreasing false positives. In a single experiment using bait Fos, more than 10 interactors, including not only direct, but also indirect interactions, were enriched. Further, previously unidentified proteins containing novel leucine zipper (L-ZIP) motifs with minimal binding sites identified by sequence alignment as functional elements were detected as a result of using a randomly primed cDNA library. Thus, we consider that this simple IVV selection system based on cell-free cotranslation could be applicable to high-throughput and comprehensive analysis of PPI and complexes in large-scale settings involving parallel bait proteins.

In vitro selection experiments using mRNA display methods such as IVV (Nemoto et al. 1997; Miyamoto-Sato at al. 2003) or mRNA-peptide fusions (Roberts and Szostak 1997; Hammond et al. 2001; Keefe and Szostak 2001), which were originally developed for evolutionary protein engineering, are expected to be powerful tools for analyses of protein functions in the post-genomic era (Amstutz at al. 2001). Although applications to both evolutionary engineering (Keefe and Szostak 2001) and genome analyses (Hammond et al. 2001) have been reported, current protocols are too technically demanding and tedious to use extensively or to automate for genome-wide experiments (Gold 2001). The results of large-scale yeast proteome analysis using the two-hybrid method (Uetz et al. 2000; Ito et al. 2001) and TAP-mass spectrometry (TAP-MS) method (Rigaut et al. 1999; Gavin et al. 2002) have been reported. These data have been analyzed intensively, and indicated that existing large-scale protein interaction data sets are nonsaturating, and that integrating many different experimental data sets can yield a better overall view than a single method alone (Bader and Hogue 2002; Legrain et al. 2002). Therefore, mRNA display should become a useful method, if it can be adapted for high-throughput in vitro analysis of protein–protein interactions (PPI) and complexes.

Basically, mRNA display is composed of four essential processes, i.e., transcription, translation, selection, and RT–PCR. Although mRNA display itself is expected to be a high-throughput technique because of the use of a library, the current protocol (Roberts and Szostak 1997; Hammond et al. 2001; Keefe and Szostak 2001) of mRNA display is tedious and requires additional post-translational processes such as isolation of mRNA–protein fusions. A stable and efficient IVV (Miyamoto-Sato at al. 2003) should allow simple selection without any requirement for post-translational processes. In addition, in vivo preparation of bait proteins (Hammond et al. 2001) is required for PPI analysis, and this would be disadvantageous for high-throughput, genome-wide analysis with multiple bait proteins. Our newly developed cell-free cotranslation technique provides a totally in vitro manipulation that is suitable for high-throughput, genome-wide analysis as a result of in vitro bait translation instead of in vivo bait preparation (Fig. 1A). Cotranslation of bait and prey proteins should also be advantageous for the formation of protein complexes (Fedorov and Baldwin 1997; Jemutus et al. 1998). This approach would offer a good chance to obtain a comprehensive data set including not only direct, but also indirect interactions in a single experiment.

A proportion of false positives arises in PPI analyses with methods such as the two-hybrid method (Legrain and Selig 2000; Legrain et al. 2001; Bader and Hogue 2002; Legrain 2002). Decreasing false positives is important to obtain data that are, at least potentially, biologically relevant. Accordingly, we used two-step purification of the IVV selection (Fig. 1A) based on the TAP method (Rigaut et al. 1999), which is suitable for analysis of protein complexes formed by capture from a crude mixture with a tagged bait protein, with a low level of false positives (Rigaut et al. 1999). To further decrease false positives and to obtain information about direct/indirect interactions, an in vitro post-selection was performed. The post-selection is composed of a pull-down assay to confirm the interactions using C-terminal protein labeling (Nemoto et al. 1999; Miyamoto-Sato et al. 2000, 2003; Doi et al. 2002) and real-time polymerase chain reaction (PCR) assay to confirm the enrichments. The use of two-step
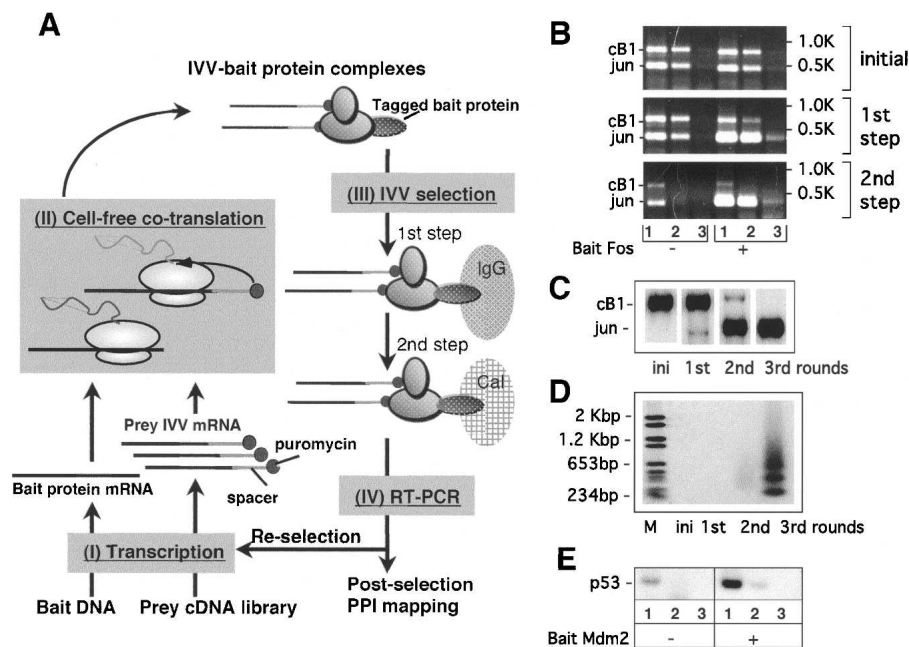
Figure 1. IVV selection based on cell-free cotranslation. (*A*) Design of the IVV selection. The selection is composed of transcription (I), cell-free cotranslation (II) coupling translation and interaction of tagged bait protein and prey IVV library to form complexes, IVV two-step selection (III) (Rigaut et al. 1999), and RT–PCR (IV) to amplify the library. (IgG) IgG beads for the protein A (z domain) tag used for the 1st step of the purification; (Cal) Calmodulin beads for the calmodulin binding protein (CBP) tag used for the 2nd step of the purification. (*B*) Enrichment in the 1st and the 2nd steps of the two-step selection in IVV. The enrichment was confirmed with (Fos+) or without (Fos−) 400 nM *fos* as bait and IVV templates of 100 nM *jun* and *cB1* as prey. (Lanes *1–3*) RT–PCR analysis with original eluate and 100-fold dilutions to 10,000-fold, respectively. (*C*) Enrichment after the two-step selection of IVV. RT–PCR analysis was performed with 100 nM *cB1* and 10^{-2} nM *jun*. (Ini) Initial mixture of *jun* and *cB1*. (*D*) Enrichment of Jun from a mouse brain cDNA library. Southern blots were done with RT–PCR products of the two-step selection of IVV using Dig DNA labeling and detection kit (Roche) with a *jun* (616–1005) template labeled with a random hexamer primer. (Ini) Initial library. (*E*) Enrichment of p53 from a human brain cDNA library. PCR (1 cycle of 98°C, 2 min; 35 cycles of 98°C, 30 sec; 62°C, 30 sec; 73°C, 30 sec; one cycle of 73°C, 2min) was done with human_p53-5′40–63 (ctgagtcaggaaa cattttcagac) and human_p53-3′257–276 (gggccaggagggggctggtg) using RT–PCR products after the 1st step of the two-step selection of IVV. The enrichment was confirmed by comparing the results with Mdm2 (+) or without Mdm2 (−) as bait. (Lanes *1–3*) PCR analysis with the original eluate and 100-fold dilutions to 10,000-fold, respectively.

(Fig. 1A). We investigated whether or not the enrichment of the in vitro selection based on cell-free cotranslation using IVV (Miyamoto-Sato et al. 2003) with Fos as the bait is sufficiently informative (Fig. 1A). Here, a two-step purification of the IVV selection was performed. After one round of the two-step purification of the IVV selection, RT–PCR indicated that Jun of AP1 complex (Fos/Jun) was enriched 1000-fold relative to cB1, a negative control (Fig. 1B, 2nd step). In the selection from a mixture of *jun* and *cB1* IVV templates (*jun*/ *cB1* = 1/10^4; Fig. 1C, ini), Jun was enriched over 10^8-fold relative to cB1 after three rounds of selection (Fig. 1C). This suggests that the average enrichment might be approximately three orders of magnitude per round, which is sufficient for in vitro selection experiments (Roberts and Szostak 1997). In fact, Jun of AP1 complex was clearly enriched after three rounds of selection in the selection from a mouse brain cDNA library (Table 1, Experiment 1; Fig. 1D). Further, Mdm2/p53 complex was enriched 100-fold relative to Mock (bait Mdm2 [−]) from a human brain cDNA library after the 1st step of two-step purification of the IVV selection (Fig. 1E), in good agreement with the result for the Fos/ Jun complex. Thus, this mRNA display using a previously developed highly stable and efficient IVV (Miyamoto-Sato et al. 2003) is expected to have good generality. This IVV selection did not require any post-translational process such as purification of fusions. Also, the IVV (Miyamoto-Sato et al. 2003) provided a completely in vitro selection without any in vivo step as a result of the use of the cell-free cotranslation technique. This simple and totally in vitro selection should be suitable for high-throughput analysis of PPI.

purification of the IVV selection followed by post-selection should provide reliable data for PPI analysis.

Here, we present a simple and totally in vitro IVV selection method (Fig. 1A) based on cotranslation of bait and prey proteins using tagged bait Fos from AP1 complex (Fos/Jun) (Chinenov and Kerppola 2001) and a prey mouse brain cDNA library. After IVV selection, we performed sequence analysis, and the sequence data were analyzed computationally to assign genes to rapidly generate a PPI map (Fig. 2). An in vitro post-selection was carried out to confirm positive interactions before PPI mapping and in silico analysis. We show that our IVV selection and post-selection system is capable of enriching not only direct, but also indirect interactions, based on in vitro experimental evidence. Further, we demonstrate the benefits of using a random primed cDNA library instead of a full-length cDNA library.

Results

### A simple and completely in vitro selection

Our IVV selection procedure is composed of cell-free cotranslation and complex selection, followed by RT–PCR amplification

Optimized IVV selection with more interactors

We examined the effect of a difference in enrichment factor per round on the number of detectable interactions (Experiments 1 and 2). In Experiment 1, a two-step purification of the IVV selection was used in all three rounds (Fig. 1C), while in Experiment 2, a one-step purification (just the first step of the two-step purification) was used until the 3rd round, followed by a two-step purification of the IVV selection at the last 4th round. Since Jun was enriched 100-fold after one-step selection (Fig. 1B) and enriched 1000-fold after two-step purification of the IVV selection (Fig. 1B,C), Experiment 2 gave the same overall enrichment (10^9-fold) as Experiment 1. However, we found that detected gene numbers in Experiment 2 are about twofold higher than those in Experiment 1 (Table 1). One-step purification of the IVV selection imposes a lower selection pressure than two-step purification, and a lower selection pressure might yield more inter-
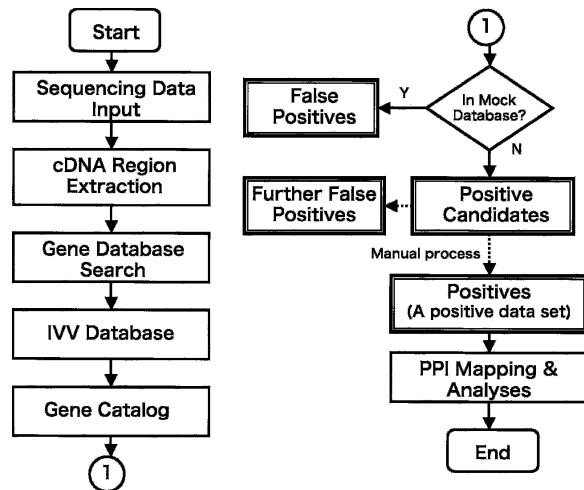
Figure 2. A flow chart of the IVV analysis system for the sequence data. The automatic sequence data analysis system provides a positive data set for PPI analysis (Fig. 5) as described in "Identification of positives for PPI mapping and in silico analysis" in the protocol section. The manual process (broken arrows) corresponds to post-selection, which consists of in vitro experiments after selection to identify independently final positives and further false positives using a real-time PCR assay (Fig. 3A) and a pull-down assay (Fig. 3B).

actors (Table 1, Experiment 2), detecting not only major, but also minor interacting proteins in a single experiment. The results suggest that the IVV selection using a two-step purification of the IVV selection for only the last round could detect more interactors (Table 1, Experiment 2). This indicates that the protocol used in Experiment 2 decreased false negatives. In the two-hybrid method, the average number of interactors was three using the matrix approach (Legrain and Selig 2000), while here we detected more than 10 interactors (Table 1, Experiment 2).

In vitro post-selection to identify positives and direct/indirect interactions

In order to decrease false positives, IVV post-selection was performed by both pull-down assay (Fig. 3A) and real-time PCR (Fig. 3B). Real-time PCR can in principle detect all positive interactions, including multiple interactions, and is therefore useful to confirm protein complexes containing direct or indirect interactions in the IVV selection. We obtained the 12 genes in Table 1 as positives (77% of initial data) after in vitro post-selection in Experiment 2. Most of the positives in Table 1, such as Jun (Chinenov and Kerppola 2001), Optn (Rezaie et al. 2002), and C130020M04Rik, were confirmed by both pull-down assay and real-time PCR (a part of the data is shown in Fig. 3A,B). Troarfips (hypothetical +1 frame-shifted Maged3 [Saburi et al. 2001] and Magphinin [Saburi et al. 2001]) gave positive results in a pull-down assay, but could not be evaluated by real-time PCR because of difficulty in the primer design (Fig. 3A,B). To verify the interaction of Troarfips, an immunoprecipitation experiment was therefore performed (Fig. 3C). Accordingly, these 10 proteins, including Troarfips, were considered to be direct interactors (Table 1, Interaction, D; 83%). On the other hand, Snapc5 (Henry et al. 1998) and Rit2 gave positive results in real-time PCR, but negative results in a pull-down assay (Fig. 3A,B), and we designated them as ID interactors (interactions including indirect interactions; Table 1, Interaction, ID; 17%). For Snapc5, a direct interaction with Jun was found in a further pull-down

assay (Fig. 3D). For Rit2, reproducible interaction in two independent experiments (Table 1, Experiments 1 and 2) and positive results in real-time PCR in both cases (Table 1, ◎) were obtained. These findings suggest that Snapc5 and Rit2 might be associated with Fos in an indirect fashion. In fact, Elf1, a member of the Ets family interacting directly with Jun and indirectly with Fos (Elf1/Fos/Jun is a well-known complex [Chinenov and Kerppola 2001]), was enriched in the course of selection and confirmed by real-time PCR (Fig. 4A). This suggests that IVV selection based on cell-free cotranslation can enrich not only direct, but also indirect interactions in a single experiment. An analysis, including indirect interactions, represents a clear advance over the current mRNA display (Hammond et al. 2001) and two-hybrid systems.

A positive data set including novel interactions with novel L-ZIP motifs

After post-selection, we obtained 12 positives as a positive data set (Table 1). Here, the number of positive data is dependent on the number of sequence analyses. In fact, analysis of 168 sequences in Experiment 2 afforded 12 positive genes (Table 1), and analysis of 525 sequences in Experiment 2 provided over twofold more positive genes (data not shown). As shown in Table 1, most of the positives had L-ZIP motifs (Chinenov and Kerppola 2001). However, positives included not only known Fos interacting proteins, such as Jun (Chinenov and Kerppola 2001), Jund1 (Chinenov and Kerppola 2001), and Psmc5 (Wang et al. 1996) (yellow squares; 25%), but also unknown Fos interacting proteins (blue and white squares; 75%). These results suggest that we do not yet know all of the interactions of Fos, which is considered to be a well-characterized protein. The percentage of interaction-known proteins was around 25% in this experiment. This is consistent with a previous study (Uetz et al. 2000; Ito et al. 2001), in which an analysis of two-hybrid interactions provided a huge number of novel interactions. Moreover, we detected function-unknown (novel and hypothetical) proteins (white square; 33%) including two hypothetical proteins with novel L-ZIP motifs (Fig. 4B). These proteins are possible products of novel alternative reading frames (hypothetical +1 frame-shifted Maged3 [Saburi et al. 2001] and Magphinin [Saburi et al. 2001]). We named them Troarfip1 (hypothetical Trophinin [Saburi et al. 2001] alternative reading frame [Kamijo et al. 1997] Fos interacting protein 1) and Troarfip2 (Table 1). We found a hypothetical 191 amino acid protein frame for Troarfip1 in Maged3 (AF319977; 758–1333) and a hypothetical 184 amino acid protein frame for Troarfip2 in Magphinin (AF331848; 5129–5683). Surprisingly, Maged3 (Saburi et al. 2001) and Magphinin (Saburi et al. 2001) themselves have no L-ZIP motifs and no interactions with Fos in vitro (data not shown); however, both +1 frame-shift Troarfips (Troarfip1 and Troarfip2) have L-ZIP motifs (Table 1) and interact directly with Fos (Fig. 3A,C). In addition, Troarfips have novel L-ZIP motifs, which we named super L-ZIP (Fig. 4B). This motif has an interesting construction composed of two different phases of L-ZIP placed alternately (Fig. 4B). The results suggest that a randomly primed cDNA library has the potential to allow the discovery of novel alternative reading frame proteins with novel functional motifs.

PPI mapping and function analysis

A PPI map (Fig. 5) for Experiment 2 (Table 1) was automatically prepared as a result of automatic sequence data analysis with IWAS (Fig. 2), and included 12 positive D interaction (solid black

Table 1. Positive data set of a gene catalog

Gene name	Experiment 1 (160 seq.)	Experiment 2 (168 seq.)	Total (328 seq.)	L-ZIP	Real-time PCR	Pull-down	Interaction	Common hit position
Jun	40	48	88	○	◎	○	D	1799···1853
Troarfip1 (+1 frame shifted Maged3)	53	19	72	○$	N.D.	○	D	1093···1249
Troarfip2 (+1 frame shifted Magphinin)	19	41	60	○$	N.D.	○	D	5499···5533
Eef1d (eukaryotic translation elongation factor 1 delta)	6	1	7	○	◎	○	D	300···560
Schip1 (Schwannomin interacting protein 1)	3	2	5	○	◎	○	D	638···757
Optn (Optineurin)		5	5	○	◎	○	D	602···789
Jund1		3	3	○	◎	○	D	977···1127
C130020M04Rik		3	3	○	◎	○	D	1281···1146
Psmc5 (mSUG1, ATPase)		2	2	○	◎	○	D	129···242
FLJ32000 (LOC362577)		2	2	○	◎	○	D	2363···2532
Snapc5 (2010203A03R)		2	2	○	◎	×	ID	3···156
Rit2 (Ras-like protein)	1	1	2	×	◎	×	ID	674···822

Experiment 1: 3 rounds of two-step selection (3rd round 2nd step, 160 sequences).
Experiment 2: 3 rounds of one-step selection followed by one round of two-step selection (4th round 2nd step, 168 sequences).

D : Direct interactions
ID: Interactions including indirect ones

$: Novel L-ZIP (Super L-ZIP)
○: Positive   ×: Negative

◎: Positives in both Experiment 1 and 2
N.D.: Not determined because of primer design difficulty

▨ : Function-known and Fos interaction-known genes

▨ : Function-known and Fos interaction-unknown genes

☐ : Function-unknown (Novel and hypothetical genes)

lines in Fig. 5; 83%) and ID interaction (dotted black lines in Fig. 5; 17%) proteins (Table 1). ID interaction should include indirect interactors such as Elf1 (Fig. 4A). The percentage of function- and Fos interaction-known genes in this experiment is just 25% (Table 1, yellow squares). To identify the functions of the other 75%, an in silico analysis was performed by using LocusLink (Wheeler et al. 2004; Fig. 5). We found that C130020M04Rik, a function-unknown gene, has annotations related to transcription-regulator activity (Fig. 5). Therefore, it is reasonable that C130020M04Rik interacts with the transcription regulation factor Fos (Fig. 3A,B). Actually, C130020M04Rik has sequence homology with the ATF family, which is closely related to the AP1 family. Thus, C130020M04Rik could be closely related to the transcription regulator AP1 family. An in silico analysis can give us a good guide to protein function if well-known proteins surround a function-unknown protein having useful annotations. The results suggest that an in silico analysis based on a PPI map could both support data credibility and be helpful in finding interesting and significant proteins using databases.

Discussion

One of the problems in biotechnological applications of cell-free translation is whether the proteins obtained can fold correctly and mutually interact. Cotranslational folding/association is an essential characteristic of prokaryotes and eukaryotes, even in cell-free translation systems (Fedorov and Baldwin 1997; Jemutus et al. 1998). Accordingly, cell-free cotranslation (coexpression), coupling translation and interaction of bait and prey proteins (Fig. 1A), should favor the formation of protein complexes. The fact that a protein interaction analysis including both direct and indirect interactions was achieved (Fig. 4A, Elf1), suggests that IVV selection based on cell-free cotranslation is an advance over both the current mRNA display (Hammond et al. 2001) and two-hybrid systems. Also, an in vitro post-selection can define direct/

indirect interactions systematically. However, it is not always possible to confirm all components of protein complexes by real-time PCR and pull-down assay, because these assays sometimes fail to detect multiple complex partners.

In contrast to the TAP-MS method, the totally in vitro manipulation of IVV would not be biased by cytotoxic or secretion-incompatible cDNA products. This is an advantage over in vivo methods. Also, IVV can detect proteins present in small amounts, because it analyzes genome sequences by using the amplification of mRNA tags instead of amino acid sequences. In this study, the use of a randomly primed cDNA library provided functional elements as minimal or common binding sequences (Table 1, Common hit position) identified by sequence alignment and resulted in the identification of novel L-ZIP motifs (Fig. 4B); thus, in principle, it allows genome analysis annotated with precisely defined binding sites. This is an advantage over the analysis of amino acid sequences by the TAP-MS method.

Regarding false positives, we used plural selection rounds and a two-step purification of the IVV selection, followed by a post-selection using both real-time PCR and pull-down assays (Figs. 1A, and 3A,B) to minimize in vitro technical or experimental nonspecific interactions (false positives). Indeed, the 12 genes in Table 1 identified as positives after post-selection appear to have been identified with high reliability (Fig. 3A–D), since the validity of most of them is supported by additional evidence, such as plural sequences, L-Zips, common hit positions, and reproducible specific enrichment, confirmed by real-time PCR, in both Experiments 1 and 2 using different selection protocols (Table 1). Thus, we consider that our totally in vitro selection using IVV is reliable for the identification of physicochemical interactions with low false positives. We believe that the aim of procedures such as IVV selection is to identify in vitro physicochemical interactions as thoroughly as possible. The biological relevance of the identified interactions must then be addressed using other methods. Nevertheless, because we do not have
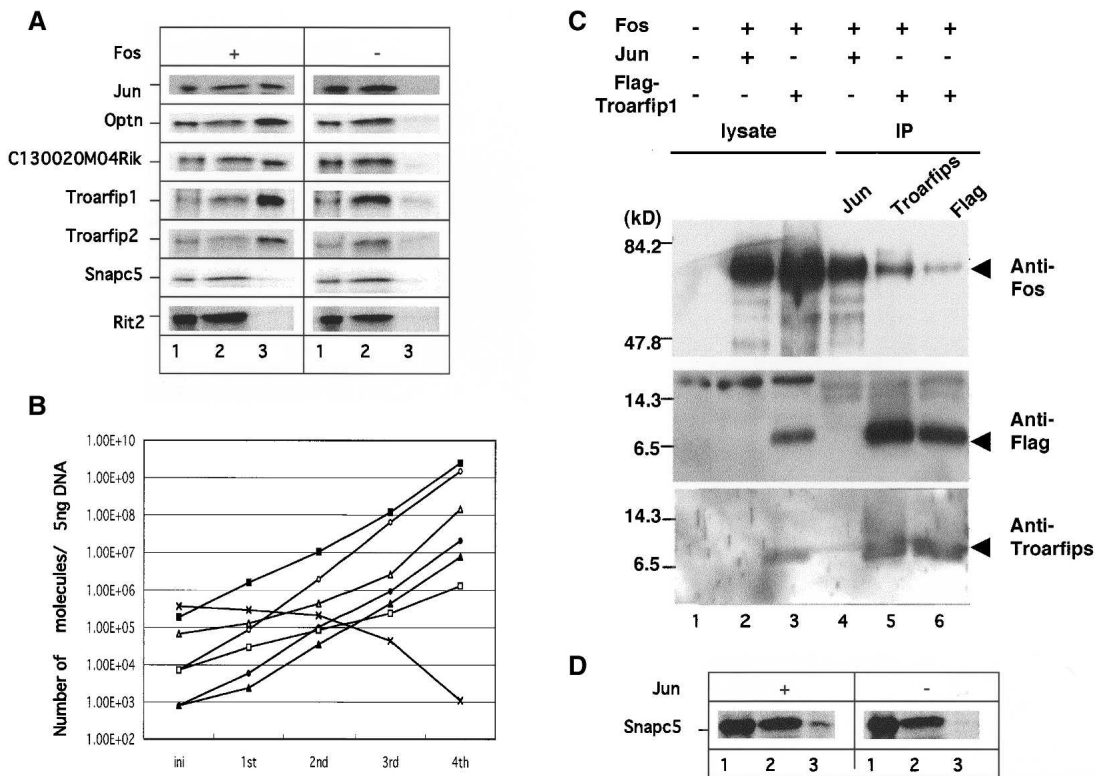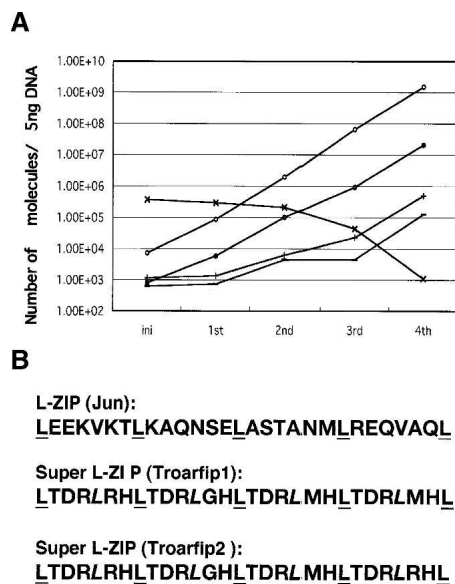
Figure 3. Post-selection and further confirmations. (*A*) Interaction analyses by pull-down assay. Pull-down assays of all 12 genes listed in Table 1 using C-terminally labeled proteins (Miyamoto-Sato et al. 2000; Doi et al. 2002) were performed with (+) or without (−) Fos as bait. The results of 1% SDS-PAGE for seven genes are shown. (Lanes *1–3*) Translation product (input), supernatant, eluate (output), respectively. (*B*) Enrichment analyses with real-time PCR. All 12 genes listed in Table 1 were analyzed in each round's pool (1st to 4th) of Experiment 2, and results for six genes, Jun (○), Jund1 (●), Rit2 (□), Optn (■), Snapc5 (△), C130020M04Rik (▲), are shown. (Ini) Initial library. β-Actin (×) is a negative control. (*C*) Further confirmation of Troarfips. Coimmunoprecipitation (IP) of Fos and Troarfip1 in COS-1 cells was performed as described in "Methods". (Lanes *1–3* [lysate]) Negative control (background), coexpression of Fos and Jun, and coexpression of Fos and Flag-Troarfip1, respectively. (Lanes *4–6* [IP]) IP with Jun using coexpression of Fos and Jun, IP with Troarfips using coexpression of Fos and Flag-Troarfip1, and IP with Flag using coexpression of Fos and Flag-Troarfip1, respectively. (*D*) Further confirmation of Snapc5. Pull-down assays of Snapc5 in Table 1 using C-terminal labeled proteins were performed with (+) or without (−) Jun as bait, followed by analysis on 15% SDS-PAGE. (Lanes *1–3*) Translation product (input), supernatant, eluate (output), respectively.

knowledge of all possible in vivo events or all interactions that may occur in normal or pathological conditions, in vitro interaction data can provide clues to uncover new pathways or mechanisms of biological functions or diseases.

There remains the problem of false negatives buried in the library. False negatives could depend on the number of sequences determined, the initial number of genes in the library, and the interacting affinity. For example, Jun, Jund1, and Junb, which are all members of the Jun family, were all detected by real-time PCR (Fig. 4A). Although Jun and Jund1 were also detected by sequence analysis, Junb was not detected, presumably because of insufficient enrichment. Nevertheless, we obtained more than 10 reliable interactors using bait Fos in a single IVV selection (Table 1, Experiment 2), whereas the average number of interactors was only three in the two hybrid method (Legrain and Selig 2000). This suggests that not only false positives, but also false negatives might be relatively lower in the IVV selection and post-selection system. However, further experiments will be needed to confirm this view. Recently, the intersection of two protein/peptide interaction data sets obtained with yeast two-hybrid and phage display strategies was utilized to extract more meaningful biological information on the binding partners (Legrain 2002). We here propose to use puromycin technology (Miyamoto-Sato et al. 2003) in conjunction with IVV selection to identify com-

plex interactions with low false positives and microarrays of C-terminally labeled proteins (Doi et al. 2002; Kawahashi et al. 2003) to identify direct interactions with low false negatives, thereby allowing the efficient extraction of physicochemical interactions and complexes.

To confirm the advantages of our method, detailed comparative studies with existing methods will be needed. It is clear that mRNA display using the IVV method would represent an advance over two-hybrid systems and the TAP method for detecting toxic proteins, due to the use of only in vitro manipulations. Further, although existing methods of mRNA display, such as two-hybrid systems, can extract binary interactions and define binding sequences, our mRNA display can detect not only binary interactions, but also protein complexes, due to the cotranslation and selection technique of the IVV method (Table 1; Figs. 1A, 3, and 4A). Our method should therefore yield new insights into protein interaction networks as a result of detecting protein complexes, as well as defining binding sequences. Our results here demonstrate that the use of plural selection rounds and a two-step purification of the IVV selection (Fig. 1A) provides a reliable procedure for picking up physicochemical interactions with a low level of false positives due to in vitro technical or experimental nonspecific interactions (Table 1). In principle, mRNA display can never avoid false negatives, because the large library size

A

B

L-ZIP (Jun):
LEEKVKT<u>L</u>KAQNSE<u>L</u>ASTANM<u>L</u>REQVAQ<u>L</u>

Super L-ZI P (Troarfip1):
<u>L</u>TDR*L*RH<u>L</u>TDR*L*GH<u>L</u>TDR*L* MH<u>L</u>TDR*L*MH<u>L</u>

Super L-ZIP (Troarfip2):
<u>L</u>TDR*L*RH<u>L</u>TDR*L*GH<u>L</u>TDR*L* MH<u>L</u>TDR*L*RH<u>L</u>

Figure 4. Detection of a well-known protein complex and novel L-ZIP motifs. (*A*) Evidence of a protein complex enrichment. The real-time PCR of Junb (+) and Elf1 (−) in each round's pool (from 1st to 4th) of Experiment 2 in Table 1. (Ini) Initial library. Jun (○) and Jund1 (●) are positive controls and β-actin (×) is a negative control. Jun, Jund1, and Junb are members of the Jun family that interact directly with Fos. Elf1 is a well-known protein, coprecipitated with Fos/Jun family complexes and interacting indirectly with Fos mediated by Jun family members (Chinenov and Kerppola 2001). (*B*) Novel L-ZIP motifs in hypothetical frame-shifted proteins. L-ZIP of Jun has a single L-ZIP, which is composed of the underlined leucines every seven amino acids. The super L-ZIP of Troarfip1 or Troarfip2 has a dual-L-ZIP, which is composed of the underlined leucines and the italic leucines every seven amino acids, alternately. Both super L-ZIP motifs are almost the same except for one amino acid.

($10^{12–13}$) is beyond the feasible capacity for sequencing analysis; nevertheless, we think that false negatives in our method might be lower than in other methods, because the number of interactors identified per selection experiment seems to be substantially larger than in existing methods, although this naturally requires further confirmation.

In this work, mRNA display using IVV has been developed as a useful tool suitable for high-throughput PPI and complexes analysis (Figs. 1, 3, and 4A) using simple and totally in vitro selection based on a cotranslation technique and a two-step purification. It is only necessary to put tagged bait and prey mRNA templates into a cell-free translation system, followed by selection (Fig. 1A). After IVV selection, the relatively laborious steps of cloning and sequencing were performed using high-throughput automated laboratory instruments, and the mass sequence data obtained were analyzed computationally to assign genes using our IVV analysis system (Fig. 2). We have focused on the Fos/Jun system for evaluation of our method. The analysis using a bait Fos revealed not only direct, but also indirect interactions (Figs. 3 and 4A). Also, after in vitro post-selection, we obtained more than 10 interactors (Table 1) mediated by precisely defined binding sites, including novel L-ZIP motifs (Fig. 4B), as a result of using a randomly primed cDNA library. Here, we used real-time PCR and pull-down assays to validate our method, but the results in Table 1 (a positive data set) suggest that positives could be simply defined as genes having plural sequences in one experiment or reproducible-specific enrichment in two or more experi-

ments. It may be possible to use this definition to decrease false positives without in vitro assays in the future, to maximize throughput. It is noteworthy that in silico analyses of the PPI map (Fig. 5) using available databases not only supported the credibility of the data, but also revealed key proteins for further in vivo analyses. To support the generality of the IVV method, we also present data showing the enrichment of Mdm2/p53 complex from a human cDNA library in IVV selection (Fig. 1E), in addition to the enrichment of the Fos/Jun complex from a mouse cDNA library in IVV selection (Fig. 1B–D). While the Fos/Jun complex is composed of Fos/Jun heterodimer binding to an asymmetric AP1 site (Chinenov and Kerppola 2001), Mdm2/p53 complex involves an enzyme-protein-like interaction, in which a relatively deep hydrophobic pocket of Mdm2 interacts with a helical region of p53 (Vassilev et al. 2004). The fact that we have confirmed the enrichment of completely different types of interactions using libraries from different species suggests that the IVV method is robust. Thus, we believe the IVV system is now ready for high-throughput analysis of PPI and complexes in large-scale settings with a number of parallel bait proteins.

Methods

Preparation of cDNA library

Randomly primed reverse transcription (RT), followed by ligation-mediated amplification (Chechik et al. 1996), was performed with a mouse brain poly(A)+ mRNA library (Clontech) or a human brain poly(A)+ mRNA library (Cosmo Bio) using a primer (TCATCGTCCTTGTAGTCAAGCTTNNNNNNNNNN). The 5′ phosphorylated dsDNA was obtained by using a SuperScriptII Double Strand cDNA Synthesis Kit (Invitrogen). Taking advantage of the fact that 5′ phosphorylated dsDNA allows the synthetic double strand adaptor (GGAATTCG and GAACAACAA CAACAACAAACAACAACAAAATGGCTAGCATGACTG
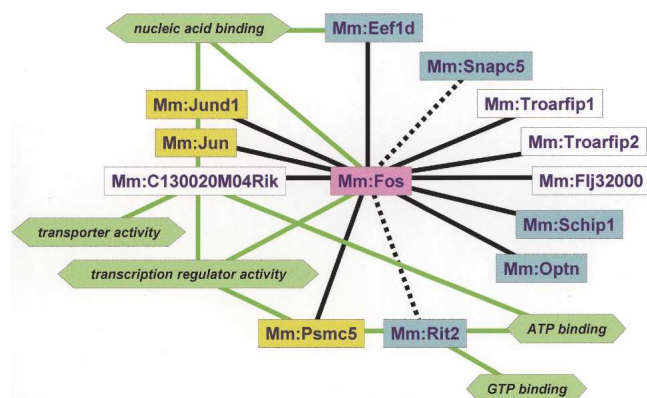
Figure 5. PPI map with function analysis. A PPI map including D (solid black lines) and ID interactions (broken black lines) of a positive data set (Table 1) was depicted using Genesphere (Fujitsu). Fos is a bait protein (pink square). Proteins (genes) having colored squares correspond to Table 1. (Yellow squares) Fos interaction-known proteins; (blue squares) Fos interaction-unknown proteins; (white squares) Function-unknown proteins. Several molecular functions (green hexagons with solid green lines) were analyzed with LocusLink. Transcription regulator activity; Fos, Jun, Jund1, Psmc5, C130020M04Rik, nucleic acid binding; Fos, Jun, Jund1, Edf1d, C130020M04Rik, ATP binding; Psmc5, Rit2, C130020M04Rik, transporter activity; C130020M04Rik, and GTP binding; Rit2. Molecular functions were defined in the gene ontologies of MGI (http://www.informatics.jax.org/searches/GO_form.shtml).

GTGGACAGCAAATGGCGAATTCC) to be ligated specifically to the 5′ end of the sense cDNA strand, the adaptor-ligated cDNA library was amplified by PCR with 5′IVVlib (GGAAGATCTATT TAGGTGACACTATAGAACAACAACAACAACAAACAACAA CAAAATG) and 3′IVVlib (tttttttttcttgtcgtcatcgtccttgtagTCAAGC). Then, transcription of the cDNA library, followed by ligation with PEG Puro spacer (Miyamoto-Sato et al. 2003), was performed (Fig. 1A, I).

Preparation of bait RNA template

The DNA template of *fos* as a bait was prepared by PCR through 15 cycles (98°C, 20 sec; 55°C, 1 min; 72°C, 4 min) after 95°C, 1 min with 5′FosCBPzz having a delta-TMV sequence (Sleat et al. 1988) (gaatttaggtgacactatagaaACAATTACTATTTACAATTACA atggctagcatgactggtggacag) and 3′FosCBPzz (GGATCTCCATTC GCCATTCA) from pCMV-FosCBPzz (Miyamoto-Sato et al. 2003), including a part of Fos (118–211), calmodulin-binding protein (CBP), and protein A (z domain), according to the TAP method[10]. Then, mRNA templates of *fos* as bait were prepared according to the previously reported method (Miyamoto-Sato et al. 2003) (Fig. 1A, I). The DNA template of *mdm2* (1–188) as a bait was prepared by cloning from a human brain poly(A)+ mRNA library (Cosmo Bio). The 1st PCR (one cycle of 94°C 2 min; 15 cycles of 94°C 30 sec, 62°C 30 sec, 73°C 2 min; 1 cycle of 73°C 15 min) was done with human_5′T7+mdm2 (atggctagcatgactggtggacagcaaatgggtcg cggatccATgTgCAATACCAACATgTCTgTAC) and human_3′mdm2– 564 (actatcagatttgtggcgttttctttgtcgttcacc), and the second PCR (same program as for the first PCR) was done with human_5′ mdm2+CBP (ggtgaacgacaaagaaaacgccacaaatctgatagtctcgagctcaa gagaagatggaaaaagaatttcatag) and 3′FosCBPzz from pCMV-FosCBPzz. Overlap PCR, using the same program, was done with human_5′T7+mdm2 and 3′FosCBPzz, and the final PCR, also with the same program, was done with 5′SP6(O′)T7 (GAATTTAG GTGACACTATAGAAacaattactatttacaattacaATGGCTAGCAT GACTGGTGGACag) and 3′FosCBPzz.

Cell-free cotranslation and two-step purification of the IVV selection

Cell-free cotranslation (Fig.1A, II) was done by using a Wheat Germ Extract (Promega) at 26 °C for 1 h with 200 nM mouse or human brain IVV mRNA library as prey and 400 nM *fos* or *mdm2* mRNA as bait in the presence or absence of 400 nM DNA (five repeats of TGACTCA sequence binding to AP1 complex of Jun/ Fos) (Chinenov and Kerppola 2001). Two-step or one-step purification (Fig.1A, III) precisely according to the TAP method (Rigaut et al. 1999) was performed as one round of IVV selection with unpurified cell-free cotranslation products, including protein complexes, captured by the tagged bait Fos. RT–PCR (Fig.1B, IV) was carried with the one-step RT–PCR kit (QIAGEN) with 5′IVVlib and 3′IVVlib to amplify the library. RT–PCR products were purified with CHROMA SPIN-1000 Columns (Clontech) and cloned with a PCR Cloning Kit (QIAGEN). Then, colony PCR was performed with Qpix (GENETIX) and a GenePCR System 2700 (Applied Biosystems) with vector-specific primers (M13R: gAAACAgCTATgACCATgATTACg and M13F: gTTTTCCCAgT CACgACgTTg). The purification of colony PCR products was done with a BioRobot 8000 (QIAGEN), and the products were sequenced with an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

Identification of positives for PPI mapping and in silico analysis

Since sequence data analysis is the most time-consuming process in IVV selection, we used automatic sequence data analysis with the IVV analysis system (IWAS) developed by Fujitsu Ltd., which is capable of dealing with 100 sequences/hour (Fig. 2). Each cDNA region is extracted by omitting vector sequences and IVV common sequences (SP6+Ω29 and T7-tag as a 5′ common sequence[2] and Flag-tag and A_8 tail as a 3′ common sequence [Miyamaoto-Sato et al. 2003]). Next, the cDNA region is analyzed using gene databases such as nt, refseq, nBLAST, and Fantom2 to obtain a gene catalog of the IVV database. To decrease false positives (Fig. 2), analyzed genes are compared with the genes of the mock database, followed by protein frame confirmation. The mock database was generated by IVV selection experiments without a bait protein. Further false positives (Fig. 2) were removed manually from positive candidates with a post-selection consisting of in vitro real-time PCR and pull-down assays. Then, a positive data set of a gene catalog (Table 1) is obtained, which is subjected to PPI mapping and in silico function analyses with LocusLink (Wheeler et al. 2004; Fig. 5) using Genesphere, used to call Xminer (Fujitsu). Genesphere (Fujitsu) is software that searches for connected and related genes and proteins from a comprehensive data set using LocusLink, UniGene, OMIM, and PubMed.

Pull-down and real-time PCR assays as a post-selection

Post-selection of real-time PCR products and pull-down assay using C-terminal protein labeling (Miyamoto-Sato et al. 2000; Doi et al. 2002) were performed completely in vitro to identify further false positives (Fig. 2). In vitro expression of all positive candidates was confirmed by C-terminal protein labeling in Wheat Germ Extract (Promega) at 26°C for 1 h using efficient mRNA templates with XA_8 as the 3′ tail instead of A_8 according to our previous report (Miyamoto-Sato et al. 2003). Pull-down assay (Doi et al. 2002) using the 1st step of the TAP method was performed with unlabeled Fos or Jun as bait protein and C-terminally labeled proteins analyzed as positives in Experiment 2 (Table 1) as prey. Preparation of *fos* as bait is described in 'Preparation of bait RNA template.' The DNA template of *jun* as a bait for pull-down assay was prepared in a similar fashion by PCR with 5′FosCBPzz and 3′FosCBPzz from pCMV-JunCBPzz, including a part of Jun (168–319) (Miyamoto-Sato et al. 2003), calmodulin-binding protein (CBP), and protein A (z domain), according to the TAP method (Rigaut et al. 1999). After cell-free translation of C-terminally labeled prey proteins (Fig. 3A,D, lane 1), C-terminally labeled prey proteins in the supernatant fraction (Fig. 3A,D, lane 2) and in the eluate fraction (Fig. 3A,D, lane 3) were obtained by pull-down assay. Analysis was done by 15% SDS-PAGE. Real-time PCR (LightCycler Quick System 330; Roche) was performed to confirm enrichment according to the product protocol with the following synthetic specific primers for Jun (Jun_F: ATCGCTCGGCTAGAGGAAA, Jun_R: CTTAAGCTGTGC CACCTGTT), Jund1 (Jund1_F: GCAAGCTGGAGCGTATCT, Jund1_R: TGACGTGGCTGAGGACTT), Optn (Optn_F: TGGG CATCGTCTCAGAAC, Optn_R: TGTGGGTGTAGGGCAGTT), Snapc5 (Snapc5_F: AAACCCTGCTGCGTCTA, Snapc5_R: ATCA TGGATTGAAGGGCTA), C130020M04Rik (C130020M04RIK_F: GGTGTCCTCCCTGGAAA, C130020M04RIK_R: TGGGCAA TCTTTATGAGCTA), Rit2 (Rit2_F: GATGATGCTTTTCAAGGCTT, Rit2_R: GGCTTTTATCTTCTTCCACA), Junb (ATCAAAGTG GAGCGAAAG, CATGACCTTCTGCTTGAG), Elf1 (Elf1_F: TGTC TAGATTGTGGGGGA, Elf1_R: GCTGACCTTCCACTTTTG), and β-actin (β-actin_F: CTTCTACAATGAGCTGCGTG,β-Actin_R: TACGACCAGAGGCATACAGG).

Antibody preparation and coimmunoprecipitation

Rabbit polyclonal antibody against Troarfips (Troarfip1 and Troarfip2) was from Immuno-Biological Laboratories (IBL) Co.,

Ltd. In brief, a synthetic peptide with the sequence GHLTDRL RHLTDRLGHLTDRLMHLTDRLMHLTDRLR was used as the anti-gen. Bovine thyroglobulin was used as the carrier protein. Im-munization was carried out eight times in total, and antiserum was collected. Since Troarfips have similar repeated sequences (Fig. 4B), it is difficult to prepare antibodies that can distinguish them. Cos-1 cells were transfected with pcDNA3.1 (Invitrogen) Fos and Flag-Troarfip1. Cloned cells were lysed in TNE buffer. Cell extracts were incubated with Protein G-Sepharose. The Sepharose was removed, then anti-Troarfips polyclonal antise-rum (IBL), anti-Flag (SIGMA), or anti-Jun (Oncogene) was added, and after incubation, Protein G-Sepharose was added. The com-plexes were washed with TNE buffer and subjected to 12% or 20% SDS-PAGE. The proteins were transferred to Immobilon-P membrane (Millipore) for Western immunoblotting with anti-Fos (Santa Cruz), anti-Flag, or anti-Troarfips.

Accession numbers

Jun; NM_010591, Troarfip1 (Hypothetical protein); +1 frame-shifted Maged3 (AF319977; 758–1333), Troarfip2 (Hypothetical protein); +1 frame-shifted Magphinin (AF331848; 5129–5683), Eef1d; NM_023240, Schip1; NM_013928, Optn; NM_181848, Jund1; NM_010592, Psmc5; NM_008950, Snapc5 (2010203A03Rik); NM_183316, C130020M04Rik; BC026483, FLJ32000 (LOC362577); XM_342896, Rit2; BC018267.

Acknowledgments

References

Amstutz, P., Forrer, P., Zahnd, C., and Plückthun, A. 2001. In vitro display technologies: Novel developments and applications. *Curr. Opin. Biotechnol.* **12:** 400–405.

Bader, G.D. and Hogue, C.W.V. 2002. Analyzing yeast protein–protein interaction data obtained from different sources. *Nat. Biotechnol.* **20:** 991–997.

Chechik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov S., and Siebert, P.D. 1996. Full-length cDNA cloning and determination of mRNA 5′ and 3′ ends by amplification of adaptor-ligated cDNA. *BioTechniques* **21:** 526–534.

Chinenov, Y. and Kerppola, T.K. 2001. Close encounters of many kinds: Fos Jun interactions that mediate transcription regulatory specificity. *Oncogene* **20:** 2438–2452.

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.

Fedorov, A.N. and Baldwin, T.O. 1997. Cotranslational protein folding. *J. Biol. Chem.* **272:** 32715–32718.

Gavin, A-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A-M., Cruciat, C-M., et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141–147.

Gold, L. 2001. mRNA display: Diversity matters during in vitro selection. *Proc. Natl. Acad. Sci.* **98:** 4824–4826.

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.

Henry, R.W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N. 1998. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes & Dev.* **12:** 2664–2672.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.* **98:** 4569–4574.

Jemutus, L., Ryabova, L.A., and Plückthun, A. 1998. Recent advances in producing and selecting functional proteins by using cell-free translation. *Curr. Opin. Biotechnol.* **9:** 534–548.

Kamijo, T., Zindy, F., Roussel, M.F., Quelle, D.E., Downing, J.R., Ashmun, R.A., Grosveld, G., and Sherr, C.J. 1997. Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product p19*ARF*. *Cell* **91:** 649–659.

Kawahashi, Y., Doi, N., Takashima, H., Tsuda, C., Oishi, Y., Oyama, R., Yonezawa, M., Miyamoto-Sato, E., and Yanagawa, H. 2003. In vitro protein microarrays for detecting protein–protein interactions: Application of a new method for fluorescence labeling of proteins. *Proteomics*. **3:** 1236–1243.

Keefe, A.D. and Szostak J.W. 2001. Functional proteins from a random-sequence library. *Nature* **410:** 715–718.

Legrain, P. 2002. Protein domain networking. *Nat. Biotechnol.* **20:** 128–129.

Legrain, P. and Selig, L. 2000. Genome-wide protein interaction maps using two-hybrid systems. *FEBS Lett.* **480:** 32–36.

Legrain, P., Wojcik, J., and Gauthier, J-M. 2001. Protein–protein interaction maps: A lead towards cellular functions. *Trends Genet.* **17:** 346–352.

Miyamoto-Sato, E., Nemoto, N., Kobayashi, K., and Yanagawa, H. 2000. Specific bonding of puromycin to full-length protein at the C-terminus. *Nucleic Acids Res.* **28:** 1176–1182.

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.

Nemoto, N., Miyamoto-Sato, E., Husimi, 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.

Nemoto, N., Miyamoto-Sato, E., and Yanagawa, H. 1999. Fluorescence labeling of the C-terminus of proteins with a puromycin analogue in cell-free translation systems. *FEBS Lett.* **462:** 43–46.

Rezaie T., Child, A., Hitchings, R., Brice, G., Miller, L., Coca-Prados, M., Heon, E., Krupin, T., Ritch, R., Kreutzer, D., et al. 2002. Adult-onset primary open-angle glaucoma caused by mutations in Optineurin. *Science* **295:** 1077–1079.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17:** 1030–1032.

Roberts, R.W. and Szostak, J.W. 1997. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci.* **94:** 12297–12302.

Saburi, S., Nadano, D., Akama, T.O., Hirama, K., Yamanouchi, K., Naito, K., Tojo, H., Tachi, C., and Fukuda, M.N. 2001. The Trophinin gene encodes a novel group of MAGE proteins, Magphinins, and regulates cell proliferation during gametogenesis in the mouse. *J. Biol. Chem.* **276:** 49378–49389.

Sleat, D.E., Gallie, D.R., Watts, J.W., Deom, C.M., Turner, P.C., Beachy, R.N., and Wilson, T.M. 1988. Selective recovery of foreign gene transcripts as virus-like particles in TMV-infected transgenic tobaccos. *Nucleic Acids Res.* **16:** 3127–3140.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403:** 623–627.

Vassliev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., et al. 2004. In vivo Activation of the p53 pathway by small-molecule Antagonists of MDM2. *Science* **303:** 844–848.

Wang, W., Chevray, P.M., and Nathans, D. 1996. Mammalian Sug1 and c-Fos in the nuclear 26S proteasome. *Proc. Natl. Acad. Sci.* **93:** 8236–8240.

Wheeler, D.L., Church, D.M., Edgar, R., Federhen, S., Helmberg, W., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., et. al. 2004. Database resources of the National Center for Biotechnology Information: Update. *Nucleic Acids Res.* **32:** D35–D40.

Web site references

http://www.informatics.jax.org/searches/GO_form.shtml; MGI.