TECHNICAL BRIEF

Compartmentalized linkage of genes encoding interacting protein pairs

Saurabh R. Nirantar and Farid J. Ghadessy

p53 Laboratory, Immunos, Singapore

Received: October 11, 2010
Revised: November 23, 2010
Accepted: December 13, 2010

Emulsion technology has been successfully applied to the fields of next-generation high-throughput sequencing, protein engineering and clinical diagnostics. Here, we extend its scope to proteomics research by developing and characterizing a method, termed *i*CLIP (in vitro compartmentalized linkage of interacting partners), which enables genes encoding interacting protein pairs to be linked in a single segment of DNA. This will facilitate archiving of the interactomes from library versus library two-hybrid screens as libraries of linked DNAs. We further demonstrate the ability to interrogate a model yeast two-hybrid *i*CLIP library for interactants by ''PCR-pulldown,'' using a primer specific to a gene of interest along with a universal primer. *i*CLIP libraries may also be subjected to high-throughput sequencing to generate interactome information. The applicability of the technique is also demonstrated in the related context of the bacterial two-hybrid system.

Keywords:
Emulsion / Interactome / In vitro compartmentalization / Technology / Yeast two-hybrid

While the number of known or predicted proteins has grown exponentially, functional annotation of these proteins has lagged behind growth in genomic data. Multiple groups have therefore carried out library versus library genome-wide screens with yeast two-hybrid [1–4], using protein chips [5] and affinity pull-downs followed by MS [6]. Among techniques commonly used to study protein interactions on a genomic level, two-hybrid systems are particularly well suited to high-throughput applications, enabling sampling of potentially all binary combinations and have contributed the majority of known interactions till date [4, 7]. However, surprisingly little overlap exists between screens carried out by different groups for the same organism [2, 3] and [1, 4], likely due to incomplete querying of all possible binary interactions as well as limitations on the number of potential positives that could be economically sequenced and false negatives inherent in the yeast two-hybrid platform [8, 9]. The solutions to these problems are to undertake more comprehensive screens, and increase the number of positive diploids that are sequenced. Such an effort is likely to place great strain on the resources of an individual laboratory.

Another consequence of the burgeoning genomic data is the absence of suitable analytical reagents such as antibodies to investigate the predicted proteome, due to the low-throughput of conventional antibody development protocols. Many groups have developed novel technologies to address this situation, involving combinatorial yeast-phage display, selectively infective phage, protein fragment complementation assays (PCAs) and selection by avidity capture (SAC) [10–13]. These techniques rely on simultaneous combinatorial expression of antigen and antibody libraries and conditional enzymatic activity or fluorescence-activated cell sorting (FACS) for selection of cognate pairs. While they are able to generate antibodies much faster than it was possible before, these methods too seem to lack simple means of storage and transmission of the antigen–antibody pairing information generated therein, and

Correspondence: Dr. Farid J. Ghadessy, p53 Laboratory, 8A Biomedical Grove, 06-06, Immunos, Singapore 138648, Singapore
E-mail: fghadessy@p53Lab.a-star.edu.sg
Fax: +65-64642049

Abbreviations: CSR, compartmentalized self replication; **iCLIP,** in vitro compartmentalized linkage of interacting partners; **SOE-PCR,** splicing by overlap extension-PCR

Colour Online: See the article online to view all figures in colour

rely on a community-wide annotation of the antigen–antibody pairs [10].

To simplify this bottleneck, Hastie and Pruitt developed a technique involving recombinase-mediated linkage of interacting genes in vivo [14]. Joining interacting genes to form a single segment of DNA naturally preserves interaction information. However, this method requires modification of existing yeast plasmid libraries and strains and is not applicable to other display systems such as yeast-phage display [10] or selection by avidity capture [13]. Furthermore, it would be desirable to preserve cognate pairing information in naturally expressed gene pairs such as antibody heavy and light chains [15]. To address these issues, we have applied emulsion technology to develop a novel, broadly applicable recombinase-independent method of linking interacting genes (called in vitro compartmentalized linkage of interacting partners, or *i*CLIP).

The method entails in vitro compartmentalization [17] of a pool of bacterial or yeast cells (selected from a library versus library interactome screen) in a modified thermostable water-in-oil emulsion. Within each aqueous compartment of the emulsion, genes encoding an interacting protein pair are linked by splicing by overlap extension-PCR (SOE-PCR) [16] (see Supporting Information Fig. S1A for schematic depiction) thereby securing interaction information. The segregation of individual cells into discrete compartments restricts non-specific linkage between cells (Fig. 1A).

Detailed materials and methods are presented in the Supporting Information section. Briefly, yeast cells (~250 000) harboring the two requisite two-hybrid vectors encoding interacting genes are added to an aqueous master mix. This comprises oligonucleotides required for linkage of cognate gene pairs by SOE-PCR, PCR buffer, a hotstart DNA polymerase, dNTPs and lyticase (to digest the cell wall). This mix is then added to an oil phase containing surfactants and stirred to form a thermostable emulsion predominantly containing one or zero cells per aqueous compartment (Fig. 1A). The emulsion is incubated at 37°C for 1 h to allow lyticase to digest the yeast cell wall, followed by a thermal cycling SOE-PCR protocol. The emulsion is next extracted into aqueous and hydrophobic phases and the linked DNA fragments from individual compartments are recovered. These can be further amplified, and subjected to a procedure we call "PCR-pulldown." This involves PCR with one gene-specific primer and one universal primer in the constant vector region to specifically amplify those SOE-PCR products containing a given gene of interest along with a gene(s) encoding interacting protein(s) (Fig. 1B). This procedure should quickly and easily allow for interactions of a given gene to be determined without having to undertake extensive and laborious screening. Another possibility is that libraries of linked genes can be subjected to high-throughput sequencing to create interaction databases or amplified and disseminated to the wider community as a resource for facile PCR-pulldown with genes of interest.

To determine the optimum emulsification conditions, cells were subjected to different stirring times using either

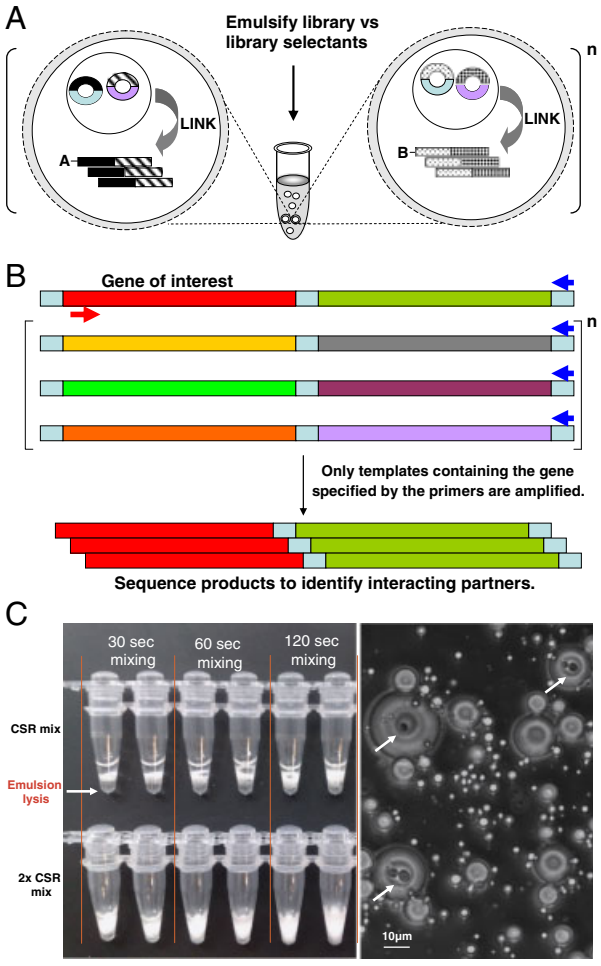

Figure 1. Schematic depiction of *i*CLIP. (A) *i*CLIP process. Positive clones from a yeast library versus library two-hybrid screen are pooled, emulsified and subjected to SOE-PCR to clonally link genes encoding interacting protein pairs. (B) The *i*CLIP library is interrogated for interacting pairs containing gene of interest. A gene-specific oligonucleotide (red arrow) along with a universal primer (blue arrow) ensures that only SOE-PCR fusions containing the gene of interest are amplified. (C) Left panel: Various mixing times and two emulsion compositions were used as indicated. Emulsion lysis was monitored by gauging the amount of aqueous phase at the bottom of the tubes (see arrow). Right panel: Brightfield images of 2 × CSR mix emulsions stirred for 120 s showing yeast cells (indicated by arrows) in individual droplets.

the previously described emulsion composition (compartmentalized self-replication (CSR) mix) [17] or a modified composition (2 × CSR mix) containing twice the concentration of surfactants. The 2 × CSR mix, stirred at 550 rpm for 120 s is optimal for the current application, as it is resistant to thermal lysis, and provides sufficiently large droplets to clonally encapsulate yeast cells (Fig. 1C). Nevertheless, a small amount of emulsion lysis by droplet coalescence does occur. This issue was dealt with by fractionating the aqueous droplets by size, into the top (T),

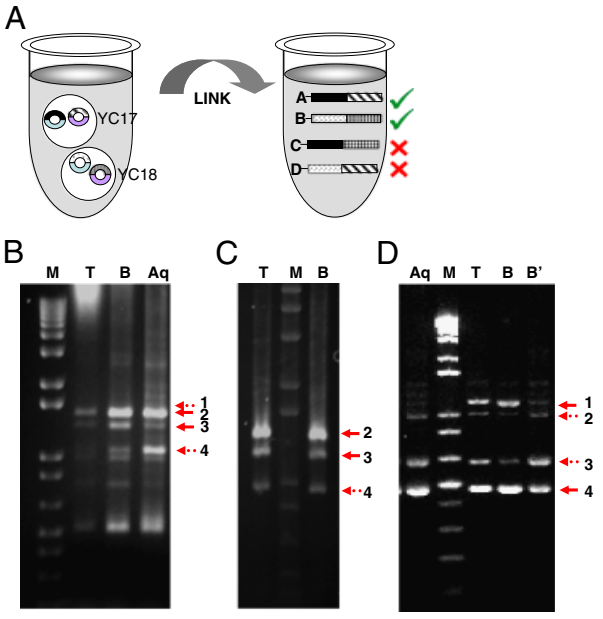

Figure 2. Proof-of-concept experiment. (A) Schematic representation of the proof-of-concept experiment. Cognate splicing of interacting genes gives rise to bands of ~1450 and 1270 bp for clones YC18 and YC17 (ticked). Non-cognate splicing leads to ~1000 and 1670 bp bands (crosses). (B) Results of the proof-of-concept experiment. Lane T: top emulsion fraction; lane B: bottom emulsion fraction; lane Aq: non-emulsified control; Lane M: DNA ladder. Correctly spliced ~1450 and 1270 bp bands arising from YC18 and YC17 are numbered 2 and 3 respectively and indicated by solid red arrows. Incorrect splicing leads to ~1000 and 1670 bp bands numbered 4 and 1 respectively (dotted arrows). The faint lower bands are unspliced individual amplicons from YC18 and YC17. Note that incorrect band 4 becomes very prominent in lane Aq when emulsion is not present; correspondingly band 3 becomes fainter. (C) Templates from lanes T and B were correctly re-amplified by emulsion PCR (D) 10^6 cells each of BC165 and BC166 were subjected to emulsion SOE-PCR followed by secondary PCR. Cognate linkage of two plasmid inserts present in BC165 and BC166 results in approximately 1150 bp (1) and 480 bp (4) bands, respectively. Incorrect (non-cognate) linkage of plasmid inserts from BC165 and BC166 results in bands of ~980 bp (2) and ~650 bp (3). Lanes T, B, Aq and M are as described in (B); Lane (B′): large-droplet fraction.

bottom (B) and large droplet (B′) fractions, separating the small intact droplets from the larger coalesced ones (see Supporting Information). Additionally, the optimal amount of lyticase required was determined to be about 170 units in 250 μL aqueous phase (Supporting Information Fig. S1).

Mis-cognate linkage during *i*CLIP can arise from either coalescence of compartments during thermal cycling, or co-compartmentalization of two or more yeast cells during emulsification. We devised a simple experiment using two characterized yeast clones YC17 and YC18 to determine the extent of mis-cognate linkage. Correct pairing results in ~1450 and 1270 bp bands; while incorrect pairing leads to bands of ~1000 and 1670 bp (see Fig. 2A for schematic explanation). The results indicate predominantly correct linkage of the respective genes (Fig. 2B). Whilst the correct-sized SOE bands predominate in the emulsified reactions (see lanes T and B), the non-emulsified control shows mostly mis-cognate linkage. Secondary PCR amplification of extracted products was carried out in emulsion [17] to prevent larger fragments being out-competed by smaller ones. The results indicate proportional amplification of the SOE products (compare intensities of correct SOE-PCR in Fig. 2B with those in Fig. 2C). Secondary amplification is particularly relevant since this enables the library to be indefinitely perpetuated and disseminated. Similar results were obtained with another pair of clones (Supporting Information Fig. S2).

Alternate two-hybrid technologies exist in *E. coli* [18] and mammalian systems [19]. Some groups have undertaken small-scale bacterial two-hybrid experiments to discover limited interaction networks [20]. We therefore carried out a proof-of-concept experiment with two bacterial clones, BC165 and BC166, similar to the experiment with yeast clones YC17 and YC18 above (details in Supporting Information). The results clearly show that emulsion SOE-PCR preserves cognate pairing, resulting in a predominance of the correctly spliced products, whereas the control reaction leads to a stochastic distribution of the various SOE-PCR product combinations (Fig. 2D). We conclude that emulsion SOE-PCR technology could also be extended to bacterial two-hybrid systems.

We next determined the *i*CLIP detection threshold, by emulsifying defined numbers of the yeast clone YC8 (harboring the human p53 gene and a potential interactant) with 250 000 cells of a test library (to simulate the representation of a given interaction in an excess of diverse unrelated interacting gene pairs). Extracted products were used as a template for PCR-pulldown with a p53-specific primer and a universal primer. Emulsification of 1000 cells but not 100 cells resulted in a clear product corresponding to a successful YC8 SOE-PCR (see Supporting Information Fig. S3). Further refinement shows that 500 target cells in a background of 250 000 library cells is the current detection threshold (Fig. 3A).

A protein of interest can be expected to interact with more than one partner. For the *i*CLIP method to be useful, it must be able to recover multiple interactions of a given protein from an excess of other interacting pairs. To simulate this situation, we took three yeast clones containing the human p53 gene, with three different potential interactants (YC8, YC12 and YC14; see Supporting Information), to represent a protein of interest with three interacting partners. The results in Fig. 3B show that the three interactions can be recovered at a dilution of 500 interaction-specific cells in 250 000 library cells as indicated by correct-sized products. A few non-specific bands can also be seen; however, these are also present in the non-emulsified aqueous controls (Fig. 3B) indicating that they arise from mis-priming events during the gene-specific PCR.

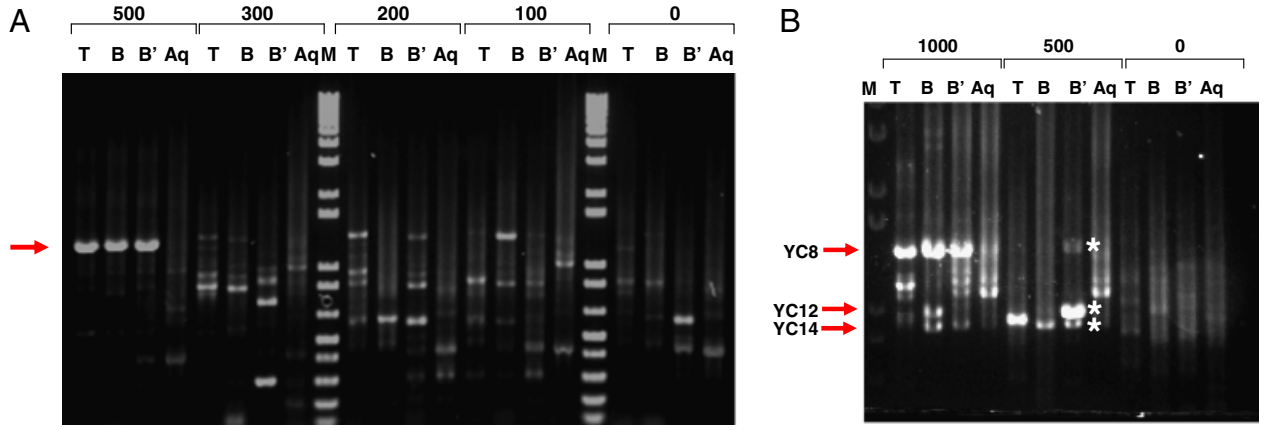

Figure 3. Threshold of detection and model selection. Lanes T, B, B′, Aq and M are as described in Fig. 2. (A) Threshold of detection. The red arrow indicates the expected SOE size. (B) 1000, 500 or 0 cells each of YC8, YC12 and YC14 were mixed with 250 000 YC16 library cells and subjected to emulsion SOE-PCR. Secondary PCR was carried out with 1 μL of the extracted products using universal primer A and gene-specific primer p53rev3 (see Supporting Information for details). The arrows indicate the expected SOE sizes (~1500 bp for YC8, ~1000 bp for YC12 and ~920 bp for YC14). Asterisks indicate the correct respective bands in the 500 cell large-droplet (B′) fraction.

The interactome library size for current genome-wide yeast two-hybrid ranges from ~1000 to about 4500 for yeast, and about 3000 for human interactome screens [1–4]. Based on our results, this would require about 2 250 000 cells to be processed for the larger of the yeast libraries. This number of cells could be readily processed in nine *i*CLIP reactions within a day. Correspondingly, about 600 reactions would be required for complete coverage of the human interactome, estimated at about 300 000 interactions [21]. While this is well within practical limits, further improvements in the linkage chemistry brought about by the use of modified oligonucleotides and/or improved polymerase enzymes will increase the efficiency of SOE-PCR (currently between 28 and 90% for different emulsion fractions, Supporting Information Fig S4) and reduce this number. A further possibility is to explore the use of alternative surfactants such as ABIL90 [22] for emulsification.

This work was funded by the Agency for Science, Technology and Research, Singapore.

The authors have declared no conflict of interest.

References

[1] Ito, T., Chiba, T., Ozawa, R., Yoshida, M. et al., A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 2001, *98*, 4569–4574.

[2] Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T. et al., Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 2005, *437*, 1173–1178.

[3] Stelzl, U., Worm, U., Lalowski, M., Haenig, C. et al., A human protein–protein interaction network: a resource for annotating the proteome. *Cell* 2005, *122*, 957–968.

[4] Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. et al., A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000, *403*, 623–627.

[5] Zhu, H., Bilgin, M., Bangham, R., Hall, D. et al., Global analysis of protein activities using proteome chips. *Science* 2001, *293*, 2101–2105.

[6] Krogan, N. J., Cagney, G., Yu, H., Zhong, G. et al., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006, *440*, 637–643.

[7] Bruckner, A., Polge, C., Lentze, N., Auerbach, D. et al., Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* 2009, *10*, 2763–2788.

[8] Huang, H., Jedynak, B. M., Bader, J. S., Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. *PLoS Comput. Biol.* 2007, *3*, e214.

[9] Ratushny, V., Golemis, E., Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system. *Biotechniques* 2008, *44*, 655–662.

[10] Bowley, D. R., Jones, T. M., Burton, D. R., Lerner, R. A., Libraries against libraries for combinatorial selection of replicating antigen-antibody pairs. *Proc. Natl. Acad. Sci. USA* 2009, *106*, 1380–1385.

[11] Jung, S., Arndt, K. M., Muller, K. M., Pluckthun, A., Selectively infective phage (SIP) technology: scope and limitations. *J. Immunol. Methods* 1999, *231*, 93–104.

[12] Secco, P., D'Agostini, E., Marzari, R., Licciulli, M. et al., Antibody library selection by the beta-lactamase protein fragment complementation assay. *Protein Eng. Des. Sel.* 2009, *22*, 149–158.

[13] de Wildt, R. M., Tomlinson, I. M., Ong, J. L., Holliger, P., Isolation of receptor-ligand pairs by capture of long-lived

multivalent interaction complexes. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 8530–8535.

[14] Hastie, A. R., Pruitt, S. C., Yeast two-hybrid interaction partner screening through in vivo Cre-mediated Binary Interaction Tag generation. *Nucleic Acids Res.* 2007, *35*, e141.

[15] Meijer, P. J., Andersen, P. S., Haahr Hansen, M., Steinaa, L. et al., Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *J. Mol. Biol.* 2006, *358*, 764–772.

[16] Horton, R. M., Ho, S. N., Pullen, J. K., Hunt, H. D. et al., Gene splicing by overlap extension. *Methods Enzymol.* 1993, *217*, 270–279.

[17] Ghadessy, F. J., Ong, J. L., Holliger, P., Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA* 2001, *98*, 4552–4557.

[18] Karimova, G., Pidoux, J., Ullmann, A., Ladant, D., A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 5752–5756.

[19] Lee, J. W., Lee, S. K., Mammalian two-hybrid assay for detecting protein–protein interactions in vivo. *Methods Mol. Biol.* 2004, *261*, 327–336.

[20] Di Lallo, G., Fagioli, M., Barionovi, D., Ghelardini, P. et al., Use of a two-hybrid assay to study the assembly of a complex multicomponent protein machinery: bacterial septosome differentiation. *Microbiology* 2003, *149*, 3353–3359.

[21] Hart, G. T., Ramani, A. K., Marcotte, E. M., How complete are current yeast and human protein-interaction networks? *Genome Biol.* 2006, *7*, 120.

[22] Ghadessy, F. J., Holliger, P., A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Eng. Des. Sel.* 2004, *17*, 201–204.

