Video Article

Identification of Protein Complexes in *Escherichia coli* using Sequential Peptide Affinity Purification in Combination with Tandem Mass Spectrometry

Mohan Babu^{1,2}, Olga Kagan¹, Hongbo Guo¹, Jack Greenblatt^{1,3}, Andrew Emili^{1,3}

¹Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto

²Department of Biochemistry, Research and Innovation Centre, University of Regina

³Department of Medical Genetics and Microbiology, University of Toronto

Correspondence to: Andrew Emili at andrew.emili@utoronto.ca

URL: <http://www.jove.com/video/4057>

DOI: [doi:10.3791/4057](https://doi.org/10.3791/4057)

Keywords: Genetics, Issue 69, Molecular Biology, Medicine, Biochemistry, Microbiology, affinity purification, *Escherichia coli*, gram-negative bacteria, cytosolic proteins, SPA-tagging, homologous recombination, mass spectrometry, protein interaction, protein complex

Date Published: 11/12/2012

Citation: Babu, M., Kagan, O., Guo, H., Greenblatt, J., Emili, A. Identification of Protein Complexes in *Escherichia coli* using Sequential Peptide Affinity Purification in Combination with Tandem Mass Spectrometry. *J. Vis. Exp.* (69), e4057, doi:10.3791/4057 (2012).

Abstract

Since most cellular processes are mediated by macromolecular assemblies, the systematic identification of protein-protein interactions (PPI) and the identification of the subunit composition of multi-protein complexes can provide insight into gene function and enhance understanding of biological systems^{1,2}. Physical interactions can be mapped with high confidence via large-scale isolation and characterization of endogenous protein complexes under near-physiological conditions based on affinity purification of chromosomally-tagged proteins in combination with mass spectrometry (APMS). This approach has been successfully applied in evolutionarily diverse organisms, including yeast, flies, worms, mammalian cells, and bacteria¹⁻⁶. In particular, we have generated a carboxy-terminal Sequential Peptide Affinity (SPA) dual tagging system for affinity-purifying native protein complexes from cultured gram-negative *Escherichia coli*, using genetically-tractable host laboratory strains that are well-suited for genome-wide investigations of the fundamental biology and conserved processes of prokaryotes^{1,2,7}. Our SPA-tagging system is analogous to the tandem affinity purification method developed originally for yeast^{8,9}, and consists of a calmodulin binding peptide (CBP) followed by the cleavage site for the highly specific *tobacco etch virus* (TEV) protease and three copies of the FLAG epitope (3X FLAG), allowing for two consecutive rounds of affinity enrichment. After cassette amplification, sequence-specific linear PCR products encoding the SPA-tag and a selectable marker are integrated and expressed in frame as carboxy-terminal fusions in a DY330 background that is induced to transiently express a highly efficient heterologous bacteriophage lambda recombination system¹⁰. Subsequent dual-step purification using calmodulin and anti-FLAG affinity beads enables the highly selective and efficient recovery of even low abundance protein complexes from large-scale cultures. Tandem mass spectrometry is then used to identify the stably co-purifying proteins with high sensitivity (low nanogram detection limits).

Here, we describe detailed step-by-step procedures we commonly use for systematic protein tagging, purification and mass spectrometry-based analysis of soluble protein complexes from *E. coli*, which can be scaled up and potentially tailored to other bacterial species, including certain opportunistic pathogens that are amenable to recombinant engineering. The resulting physical interactions can often reveal interesting unexpected components and connections suggesting novel mechanistic links. Integration of the PPI data with alternate molecular association data such as genetic (gene-gene) interactions and genomic-context (GC) predictions can facilitate elucidation of the global molecular organization of multi-protein complexes within biological pathways. The networks generated for *E. coli* can be used to gain insight into the functional architecture of orthologous gene products in other microbes for which functional annotations are currently lacking.

Video Link

The video component of this article can be found at <http://www.jove.com/video/4057/>

Protocol

1. Construction of Gene-specific SPA-tagging in *E. coli* DY330 Strain

1. The plasmid pJL148 encompassing the SPA-tag DNA sequence and the kanamycin antibiotic resistance marker cassette (Kan^R) are used as a template in polymerase chain reaction (PCR) amplification⁷. A 45 nt gene-specific forward primer, located immediately upstream of the target gene stop codon in frame with a 27 bp (5'- AGCTGGAGGATCCATGGAAAAGAGAAG -3') tag specific forward primer, and a 45 nt gene-specific reverse primer, located immediately downstream of the target gene stop codon in frame with a 27 bp (5'- GGCCCCATATGAATATCCTCCTTAGTT -3') tag specific reverse primer, are used to amplify the SPA-tag and Kan^R cassette from the pJL148 using the following PCR cycling conditions: 94 °C for 5 min; 30 cycles of 94 °C for 1 min, 55 °C for 1 min, 68 °C for 2 min 10 sec, followed by

- 68 °C for 10 min. These amplified sequences serve as substrates for the ectopically introduced λ -Red homologous recombination machinery (see below).
- The PCR product is purified using a Qiaquick PCR purification kit to remove salts that may interfere with the electroporation. The purified PCR product is subsequently targeted to integrate at the 3' end (immediately upstream of the native stop codon) of a specific gene in DY330 strain in which the λ -Red recombination machinery is expressed¹⁰.
 - The DY330 λ -Red expressing strain is grown overnight in 2 ml Luria-Bertani (LB) medium at 32 °C by shaking at 180 rpm. 1 ml of the overnight culture is subsequently inoculated into 70 ml fresh LB medium in 500 ml conical flask. The inoculum is grown at 32 °C by shaking at 180 rpm until the OD₆₀₀ reaches ~0.8.
 - The culture is transferred into a fresh 250 ml conical flask, where the cells are induced by incubating the flask in a water bath at 42 °C by gently shaking at 180 rpm for 15 min. Immediately after the induction, the flask is incubated in an ice water slurry bath for at least 30 min with shaking.
 - The ice-cold culture is immediately transferred into pre-chilled 50 ml polypropylene tube and subjected to centrifugation at 3,993 x g for 6 min at 4 °C. The cell pellet is resuspended in 50 ml ice-cold sterile water and centrifuged once again at 3,993 x g for 6 min at 4 °C. The cell pellet is then resuspended in 1 ml of cold water and transferred to a 1.5 ml Eppendorf tube, and centrifuged at maximum speed for 20 sec at 4 °C. After two or three washing steps with ice cold water, the cell pellet is resuspended finally in 700 μ l of ice-cold sterile water, resulting in electro-competent cells.
 - Through electroporation, one μ l of the purified amplicon is introduced into 40 μ l of the electro-competent cells. The cells are electroporated in a Bio-Rad GenePulser II with following settings: 2.5 kV, 25 μ F with pulse controller of 200 Ω . After homologous recombination and integration, the transformants that successfully recombined the tag/cassette into the chromosome are selected based on resistance to Kan (**Figure 1A**). Multiple transformants are selected for Western blotting to verify the correct generation (*i.e.* positive signal) of SPA-tagged fusion proteins with anti-FLAG M2 antibody that is selective against the FLAG epitopes of the SPA-tag.
 - Confirmed carboxy terminal SPA-tag fusion strain is cultured on a large-scale to isolate soluble protein complexes from harvested cells using the SPA purification protocol. The outline of the steps involved in the affinity tag purification procedure is shown in **Figure 1B**.

2. Culturing and Sonication

- Inoculate 100 μ l of a SPA-tagged *E. coli* glycerol stock into 50 ml terrific broth (TB) liquid medium supplemented with 50 μ l of 25 μ g/ml of Kan solution in a 250 ml conical flask. Grow the culture overnight at 32 °C.

Note: Since the temperature-inducible λ Red cassette in strain DY330 is under the control of a temperature-sensitive repressor¹⁰, the SPA-tagged *E. coli* strain is grown at 32 °C.

- Transfer 10 ml of the overnight culture into 990 ml fresh TB supplemented with 25 μ g/ml of Kan in a 4 liter flask. The culture is grown at 32 °C with constant shaking at 250 rpm, for 5 to 6 hr, until the OD₆₀₀ reaches to ~ 2 to 3.
- Transfer the 1 liter *E. coli* SPA-tag culture to clean centrifugation bottles and spin the cells at 4 °C in Beckman J6-HC centrifuge @ 3993 x g for 15 min.
- The supernatant is removed from the centrifugation bottles. Resuspend the *E. coli* cell pellets with 25 ml of sonication buffer. Ensure to keep the centrifugation bottles on ice at all times.
- The resuspended pellet is transferred to 50 ml polypropylene Falcon tube and frozen using liquid nitrogen. The frozen cells are stored at -80 °C for long-term use.
- Prior to sonication, thaw the frozen cells completely by keeping the pellets on ice. The thawed samples are transferred to a sterile stainless steel cup placed on ice for sonication. The probe is submerged into the sample and sonicated for 3 min. An additional 2 min is allowed to cool the sample from overheating.
- The sonicated cell lysate is transferred into the pre-chilled sterile centrifugation tube, and subjected to centrifugation at 4 °C in Beckman centrifuge using a JA-17 rotor @ 35,267 x g (16,000 rpm) for 15 min twice.

Note: In the case of membrane protein purification, the sonicated cell lysate is centrifuged only once for 15 min because we do not know in which fraction the membrane proteins are, that is, either in the pellet or in the supernatant fraction. So to avoid excess loss of membrane proteins, we spin the cells for 15 min at high-speed centrifugation and the supernatant is subsequently processed for purification (see steps below) where 1% detergent is used to solubilize the membrane proteins.

- The supernatant from the centrifugation tube is carefully transferred to a 50 ml polypropylene Falcon tube and frozen using liquid nitrogen. The sonicated frozen cell extract is stored for a maximum period of 6 months at -80 °C for future use.

3. Affinity Purification

- Prior to use, 100 μ l of anti-flag M2 beads are washed by 10 volumes of AFC buffer (30 mM Tris-HCl, 150 mM NaCl, 0.1% detergent, and 0.1-0.5 mM TCEP [tris (2-carboxyethyl) phosphine-hydrochloric acid]) without the detergent and the reducing agent TCEP (tris (2-carboxyethyl) phosphine).
- The frozen, sonicated cell extract is thawed by placing the tube in cold water. The thawed cell extract is incubated with 3 μ l of benzonase nuclease for 30 min at 4 °C. To this mixture, add non-ionic detergent Triton X-100 (final concentration of the detergent should be 0.1%) and 200 μ l suspensions of anti-flag M2 agarose beads.

Note: For all soluble protein purifications, 0.1% Triton X-100 is used to minimize non-specific adsorption, where as for purifying a membrane protein, different mild non-ionic detergents, such as C12E8 (Octaethylene glycol dodecyl ether), DDM (n-Dodecyl β -D-maltoside) and Maltose-neopentyl glycol (MNG) can be used at a 1% detergent concentration to enhance solubilization. Since different detergents have varying solubilization efficiency of membrane proteins, it is recommended to perform independent protein purifications using more than one detergent.

- The content is gently mixed by rotating the tube for 3 hr at 4 °C using a LabQuake shaker. After 3 hr of rotation, the tube is centrifuged at 1,700 x g for 6 min. The supernatant is then removed carefully, as much as possible, without disturbing the loose bead pellet.

4. The pellet is resuspended in the remaining supernatant and then transferred into a 0.8 x 4 cm Bio-Rad polypropylene prep column. Ensure to remove the bottom outlet plugs of the column, to allow the eluates to drain by gravity flow. Wash the column 5 times with TEV cleavage step.
5. After draining the washed eluates, the bottom outlet of the column is closed. To the same column, cleavage is performed by adding ~5 to 10 μ l (50 units) of TEV protease and 400 μ l of 1X AFC buffer onto the column. Ensure to close the top of the column with a cap. The column containing the beads are rotated gently overnight at 4 °C overnight using a LabQuake shaker.
6. Remove the cap on the top and the outlet plug at the bottom of the column, and drain the eluates recovered after TEV cleavage into a fresh column containing 200 μ l suspensions of calmodulin-sepharose beads, which is washed by 10 volumes of calmodulin binding buffer.
7. Both the top and the bottom of the column are fastened tightly, and the contents in the column are mixed gently by rotating for 3 hr at 4 °C using a LabQuake shaker.
8. After 3 hr of rotation, the eluate is drained by removing the top cap and the bottom plug of the column. The column is washed four times with 200 μ l of 1X calmodulin binding buffer followed by one stringent wash with 400 μ l of calmodulin wash buffer.
9. The bound protein is eluted in four fractions of 50 μ l in a fresh Eppendorf tube using 1X calmodulin elution buffer. The eluted fraction is distributed in two clean Eppendorf tubes in equal volumes. Both the tubes are subsequently dried using a speed vacuum.
10. The dried eluate from one tube is used for running a silver staining gel, while the other is stored in -80 °C, prior to using mass spectrometry.

4. Silver Staining

1. For silver staining, half the volume of 3X SDS (sodium dodecyl sulfate) sample buffer is added to 50 μ l of the dried eluate. After boiling the mixture for 5 min, the samples are loaded onto a SDS polyacrylamide gel.
2. After the gel has finished running, it is carefully placed into fixing solution (50% methanol and 10% acetic acid) and agitated gently on a rotary shaker for 20 min. The gel is then rinsed in 20% ethanol for 10 min.
3. The fixed gel is washed twice thoroughly with 500 ml of double distilled water for 10 min to ensure low uniform background.
4. The gel is then agitated gently in 500 ml sodium-thiosulfate for 1 min, followed by two washing steps with distilled water for 20 sec.
5. The water is discarded, and the gel is incubated in 200 ml of 0.1% silver nitrate for 30 min. After that time, the silver nitrate is removed and the gel is washed with distilled water for 20 sec to remove the excess silver nitrate.
6. The gel is finally hand agitated in 75 ml of the developing solution. When the desired intensity is achieved, the developing solution is discarded and 80 ml of acetic acid is added to stop the reaction. Ensure to incubate the gel in acetic acid for a minimum of 20 min for proper visualization of gel bands.

5. Proteolysis and Sample Preparation for Mass Spectrometry

1. To the dried sample, add 50 μ l of the digestion buffer and 0.9 μ l of 100 mM TCEP-HCl (tris (2-carboxyethyl) phosphine-hydrochloric acid) and incubate the mixture for 45 min at room temperature for the reduction step. Next, 1 μ l of 500 mM iodoacetamide is added and incubated in dark for another 40 min to allow for sample alkylation.
2. After the second round of incubation, add 1 μ g of immobilized trypsin to the mixture and incubate either at 37 °C for 5 hr or overnight at room temperature. Ensure to stop the reaction by adding 1 μ l of acetic acid.
3. Pre-wet the Millipore Zip-Tip pipette tip by aspirating 10 μ l of the wetting and equilibration solution. Dispense the solution to waste. Subsequently aspirate 10 μ l of the washing solution and dispense the solution to waste. Repeat this step twice.
4. For efficient binding of the peptide mixture to the tip, pipette mix the peptide mixture 20 times. The peptide mixture that adhered to the tip is washed off by aspirating and dispensing the washing solution. Repeat this procedure twice for efficient binding of the peptide mixture to the tip.
5. With the tip containing the bound peptide, aspirate 10 μ l of the wetting and equilibration solution, and dispense into a clean eppendorf tube. Repeat this step two times. After drying the eluted samples in speed vacuum, the samples can be analyzed immediately by mass spectrometry or stored at -80 °C prior to use.

6. Protein Identification by LTQ Orbitrap Velos Mass Spectrometer

The polypeptide components of the isolated complexes are identified using an LTQ Orbitrap Velos hybrid tandem mass spectrometer. The Orbitrap has exceptional resolving power (>60,000 Full Width Half Maximum, or FWHM) and mass accuracy (<2 ppm) that minimizes MS/MS sampling of irrelevant non-specific background contaminants detected in control purifications, while the high speed Velos ion trap component can detect and fragment low abundance peptides using both electron transfer dissociation and collision-induced dissociation modes. High confidence matches among resulting MS/MS spectra are mapped to reference *E. coli* protein sequences using database search algorithm like SEQUEST and each matching sequence evaluated during a probability algorithm like STATQUEST¹¹. The total spectra number, peptide sequence uniqueness and common background contaminants detected in negative control (i.e. mock) purifications are considered to achieve a low empirical false-discovery rate. The following steps are performed for protein identification:

1. The micro-columns are packed with ~10 cm of 3 μ m Luna-C₁₈ resin and are interfaced to a Proxeon nanoelectrospray ion source that is placed in line with the Orbitrap instrument.
2. A Proxeon nano flow binary HPLC pump is used to deliver a stable tip flow rate of ~300 nl min⁻¹ during the peptide separations.
3. To achieve peptide elution, an organic buffer gradient is set up according to the sample complexity. For example, the following gradient is typically set up for *E. coli* SPA samples: solvent B is increased from 2% to 6% in 1 min, to 24% in 38 min, to 100% in next 4 min, held at 100% for 1 min, then decreased to 2% in 1 min, and final hold at 2% for 15 min. The mobile phase solvent A has 95% HPLC gradient water and 5% ACN with 0.1% formic acid, while solvent B has 5% HPLC gradient water and 95% ACN with 0.1% formic acid. The flow rate at tip of the needle is set to 300 nl min⁻¹ for 60 min.

4. As the mass spectrometer cycles run through one full mass scan at 60,000 resolutions, 10 concomitant tandem fragmentation mass scans are collected for the most intense precursor ions. Dynamic exclusion, monoisotopic mass selection, and real time data-dependent acquisition instrument features are enabled.
5. The spectra are searched using a database search algorithm such as SEQUEST against a database of *E. coli* protein sequences, and the result is statistically filtered using a probability algorithm such as STAUQUEST¹¹ to ensure a low false discovery rate. Only high confidence (99% probability) candidates are selected, while matches showing more than 10 ppm mass error compared to the theoretical peptide mass are eliminated from further consideration.

Representative Results

Once tagged bait proteins, which are expressed at endogenous levels are affinity-purified from logarithmic phase cultures the samples were run on a silver-stain gel to visualize the individual polypeptide components of the isolated stable complexes. We also subjected a second portion of the affinity-purified protein samples to gel-free tandem mass spectrometry (LCMS) to identify the corresponding polypeptide sequences. The effectiveness of this APMS procedure is shown with a representative SDS-PAGE analysis of the components of RNA polymerase (RNAP) and SufBCD iron-sulfur (Fe-S) protein complexes that were affinity-purified from *E. coli* (**Figure 1C and D**). Both SPA-tagged RNAP σ^{70} (RpoD) and YacL, a protein of unknown function, co-purified specifically with the core RNAP enzyme *e.g.*, α (RpoA), β (RpoB), and β' (RpoC) subunits, and with the RNAP recycling factor (HepA). In contrast to tagged RpoD, YacL bound additionally to an essential transcription termination/anti-termination factor (NusG), suggesting a specialized function in transcription. We also detected several other smaller co-purifying proteins by LCMS that were not apparent on the gel, including the RNAP ω subunit, RpoZ, and the transcription termination/anti-termination factors, NusA and NusD, with tagged RpoD or YacL proteins, respectively. Conversely, in an independent experiment, tagged SufB, SufC, and SufD co-purified with each other, indicating joint participation in Fe-S cluster biosynthesis as single scaffolding complex. This representative example highlights the fact that previously well-studied, highly annotated bacterial multi-protein complexes participating in essential biological processes often have novel associated components that can be efficiently identified using this approach.

To define the composition of stable multi-protein complexes, we assign co-purification scores² to high-confidence interactions by taking into account the uniqueness of bait-prey, bait-bait and prey-prey relationships. Then using graph clustering procedures like Markov Clustering Algorithm², we identify discrete protein clusters from the partitioned probabilistic PPI network. Putative interaction networks can be visualized using Cytoscape¹².

Although the proteomics approach described above can reliably predict physical interactions among bacterial proteins, it may not necessarily reveal the actual functional relationship. Thus, combining proteomics data with additional functional association evidence inferred by genomics methods can be used to investigate novel mechanistic roles of previously unannotated *E. coli* proteins (**Figure 2**). For example, using an integrative machine learning approach, we have defined sub-networks of protein complexes and functional modules that participate as broader functional neighborhoods in *E. coli*².

Figure 1. Schematic overview of gene-specific SPA-tagging by recombineering and the tandem protein affinity purification procedure. Panel A: A drug selectable marker tagging cassette is amplified by the PCR using primers (1, 2) homologous to the sequences flanking on either side of a target gene's translational termination codon (3). The cassette is transformed and integrated into the *E. coli* chromosome using a transiently induced ectopic λ -Red recombination-based system. SPA, Sequential Peptide Affinity system; Kan, Kanamycin resistance cassette; *, stop codon. The figure is adapted from Babu *et al.*¹⁴. Panel B: Overview of the SPA-tag tandem protein affinity purification procedure. For membrane protein purifications, detergents are included in all buffer solutions except the final wash and elution buffers to avoid interferences with the mass spectrometry detection. Panel C: Silver-stained SDS polyacrylamide gel portraying components of SPA-tagged RNA polymerase (RpoD), a protein of unknown function (YacL) and Fe-S scaffold (SufBCD) proteins from *E. coli* after affinity purification. Panel D: Identification of RNAP and transcription/anti-termination factors co-purified from the endogenous affinity tagged RpoD (panel I) and YacL (panel II) proteins by tandem mass spectrometry. Other proteins that co-purified with tagged RpoD and YacL are not shown. [Click here to view larger figure.](#)

Figure 2. Putative *E. coli* functional neighborhoods showing assignments of cytosolic proteins of previously unknown function (from Hu *et al.*²; appended). The main hierarchical 'clustergram' (orange) shows the patterns of functional predictions and existing annotations (x-axis) for the functionally-orphan and characterized genes of *E. coli* (y-axis). Yellow and blue colors represent existing (curated databases) and inferred functions (Hu *et al.*², study), respectively, while shade intensities reflect confidence scores. Biological processes associated with various neighborhoods are indicated on the right. Insets (aquamarine, pink) show the individual components of a representative neighborhood based on the integrated physical (SPA-tagging) and functional interaction (comparative GC methods) network similarity scores. [Click here to view larger figure.](#)

Discussion

A key aspect of the SPA-based APMS approach described here is that tagging is performed within the natural chromosomal context, thereby ensuring normal gene regulation is maintained (*i.e.* native bait promoter preserved, hence expression levels is not perturbed) and native stably-associated protein complexes are recovered at near-endogenous levels. Operon polarity issues are also avoided by including an outwardly oriented promoter in the selectable marker. This SPA-tagging approach is effective enough to purify the components of low-abundance complexes to near homogeneity, including membrane-associated assemblies, even if the subunits are expressed down to only a few molecules per bacterial cell. Overall, this protocol can be easily scaled-up for purifying large-sets of tagged soluble proteins in *E. coli* or in principle any other bacterial species for which tagging via recombinering is possible, or the ones that possess naturally high transformation capability, like *Acinetobacter*, an emerging workhorse of bacterial genetics, that has been used, for example, to build a strain library of targeted gene knockouts¹³.

A major concern inherent to large-scale proteomic approaches is the identification of promiscuous non-specific interactors and spurious trace contamination. Despite two rounds of enrichment, our routine SPA purifications regularly detect frequent contaminants which are usually high-abundance housekeeping proteins, such as ribosomal proteins and chaperones, which bind non-specifically to the chromatographic resin and/or bait proteins. This problem can be mitigated in two ways. First, the total spectral number and uniqueness of each protein identified with a particular bait is considered relative to a broader set of purifications of multiple unrelated protein baits and from untagged (*i.e.* wild type) negative control strains (*i.e.* mock affinity purification experiments) to minimize false positive associations. Second, all candidate protein interactions should be validated using reciprocal tagging and purification of the putative binding partner, with the goal of independently confirming individual protein-protein interactions to avoid rare cases in which the presence of the SPA-tag perturbs incorporation of native proteins into the endogenous protein assembly.

APMS-based proteomic surveys are limited by the lack of sensitivity in terms of identifying transient or sub-stoichiometric interactions. In such cases, single-step purification procedures may be used to enhance detection of weaker interaction partners. During the course of our ongoing, decade long study of the *E. coli* interactome, we have seen the detection capabilities of new generation mass spectrometers improve steadily. Older instrumentation is biased to recurrent identification of highly abundant proteins, often precluding detection of less abundant but potentially very important interacting proteins. Conversely, the Orbitrap Velos hybrid mass spectrometer we currently use has markedly improved resolution, mass accuracy, and sensitivity thereby allowing far more efficient, high throughput detection of low abundance proteins, including transient interactions. Finally, if the nature of the interaction is completely unknown we suggest researchers to: (i) apply stringent criteria to

assign confidence scores to all putative database search matches. Proteins with two or more unique peptides are usually considered positive, assuming each match passes with a minimum likelihood threshold cut-off of 99% or greater probability; (ii) make sure to examine the specificity of candidate protein interactors recorded with the tagged bait in comparison to results obtained with negative control untagged strains (mock experiments) or unrelated protein baits; (iii) reciprocally confirm potential interactions containing an unknown protein of interest by creating the corresponding SPA tag bait fusion for large-scale purification or validate pair-wise interactions using traditional co-immunoprecipitation with protein specific or tag specific antibodies.

To date, using this systematic approach, we have managed to tag and purify around two-thirds of the ~2,750 *E. coli* soluble proteins that detectably expressed under culture in rich medium^{1,2}. We have also adapted this same basic method to systematically isolate membrane-associated protein complexes, and are now attempting to purify each of the remaining ~1,000 *E. coli* proteins predicted to be membrane bound. While the purification of membrane proteins often poses unique challenges because they are often not efficiently solubilized by the extraction buffer that we normally use for the SPA method, the addition of non-ionic detergents to our buffers enables the solubilization and purification of a majority of the *E. coli* membrane proteins we have attempted to date (Babu *et al.*, unpublished data).

Disclosures

No conflicts of interest declared.

Acknowledgements

This work was supported by funds from the Canadian Foundation for Innovation, Genome Canada, the Ontario Genomics Institute, the Ontario Ministry of Innovation, and the Canadian Institutes of Health Research grant to J.G. and A.E. The Red-expressing *E. coli* strain DY330 was a kind gift from Donald L. Court (National Cancer Institute, Frederick, MD).

References

1. Butland, G., *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*. **433**, 531-7 (2005).
2. Hu, P., Janga, S.C., Babu, M., Diaz-Mejia, J.J., & Butland, G. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* **7**, e1000096 (2009).
3. Krogan, N.J., *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*. **440**, 637-43 (2006).
4. Mak, A.B., *et al.* A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol. Cell Proteomics*. **9**, 811-23 (2010).
5. Polanowska, J., *et al.* Tandem immunoaffinity purification of protein complexes from *Caenorhabditis elegans*. *Biotechniques*. **36**, 778-80, 782 (2004).
6. Veraksa, A., Bauer, A., & Artavanis-Tsakonas, S. Analyzing protein complexes in *Drosophila* with tandem affinity purification-mass spectrometry. *Dev. Dyn.* **232**, 827-34 (2005).
7. Zeghouf, M., *et al.* Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.* **3**, 463-8 (2004).
8. Puig, O., *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*. **24**, 218-29 (2001).
9. Rigaut, G., *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-2 (1999).
10. Yu, D., *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978-83 (2000).
11. Kislinger, T., *et al.* PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol. Cell Proteomics*. **2**, 96-106 (2003).
12. Vlasblom, J., *et al.* GenePro: a Cytoscape plug-in for advanced visualization and analysis of interaction networks. *Bioinformatics*. **22**, 2178-9 (2006).
13. de Berardinis, V., *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
14. Babu, M., *et al.* Sequential peptide affinity purification system for the systematic isolation and identification of protein complexes from *Escherichia coli*. *Methods Mol. Biol.* **564**, 373-400 (2009).