research articles Journal of proteome research

A General System for Studying Protein−Protein Interactions in Gram-Negative Bacteria

Dale A. Pelletier,*,†,+ Gregory B. Hurst,‡,+ Linda J. Foote,† Patricia K. Lankford,† Catherine K. McKeown,† Tse-Yuan Lu,† Denise D. Schmoyer,§ Manesh B. Shah,† W. Judson Hervey IV,‡,◆ W. Hayes McDonald,‡ Brian S. Hooker,⊥ William R. Cannon,▽ Don S. Daly,▽ Jason M. Gilmore,▽ H. Steven Wiley,○ Deanna L. Auberry,⊥ Yisong Wang,† Frank W. Larimer,† Stephen J. Kennel,†,¶ Mitchel J. Doktycz,† Jennifer L. Morrell-Falvey,† Elizabeth T. Owens,† and Michelle V. Buchanan||

Biosciences Division, Chemical Sciences Division, Computer Science and Mathematics Division, and Physical Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, Biological Sciences Division, Computational Sciences and Mathematics Division, Environmental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, and Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

Received March 11, 2008

One of the most promising methods for large-scale studies of protein interactions is isolation of an affinity-tagged protein with its *in vivo* interaction partners, followed by mass spectrometric identification of the copurified proteins. Previous studies have generated affinity-tagged proteins using genetic tools or cloning systems that are specific to a particular organism. To enable protein−protein interaction studies across a wider range of Gram-negative bacteria, we have developed a methodology based on expression of affinity-tagged "bait" proteins from a medium copy-number plasmid. This construct is based on a broad-host-range vector backbone (pBBR1MCS5). The vector has been modified to incorporate the Gateway DEST vector recombination region, to facilitate cloning and expression of fusion proteins bearing a variety of affinity, fluorescent, or other tags. We demonstrate this methodology by characterizing interactions among subunits of the DNA-dependent RNA polymerase complex in two metabolically versatile Gram-negative microbial species of environmental interest, *Rhodopseudomonas palustris* CGA010 and *Shewanella oneidensis MR-1*. Results compared favorably with those for both plasmid and chromosomally encoded affinity-tagged fusion proteins expressed in a model organism, *Escherichia coli*.

Keywords: protein interactions • protein complexes • RNA polymerase • affinity purification • mass spectrometry

Introduction

Most polypeptide chains function as components of larger protein complexes or "molecular machines".[1–3] Living organisms gain several advantages from a strategy of assembling machines from smaller components, such as a smaller probability of translation error for small versus large proteins, the increased flexibility and coding economy when using a single subunit protein in several different complexes, more efficient channeling of enzyme substrates, and the ability to modulate the activity or function of a complex by altering its subunit composition.[1–4] Discovering the protein−protein interactions that make up these complexes is an important technique for studying protein function. If a previously unknown protein is observed to interact with other proteins or a complex of known function, then one has gained insight into the function of the unknown.[5] Genome-wide elucidation of protein−protein interactions yields a network structure that may result in higher-level information about interactions within and among protein complexes.[6–11]

Various approaches allow identification of physical protein−protein interactions in complexes.[3,12,13] Classical "one-at-a-time" experimental approaches focus on proteins that can be linked with a measured activity. Newer techniques seek to

* To whom correspondence should be addressed. E-mail: pelletierda@ornl.gov. Phone: 865-576-2857. Fax: 865-574-6210.
† Biosciences Division, Oak Ridge National Laboratory.
+ These two authors contributed equally to the writing of this manuscript.
‡ Chemical Sciences Division, Oak Ridge National Laboratory.
§ Computer Science and Mathematics Division, Oak Ridge National Laboratory.
◆ Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory.
⊥ Biological Sciences Division, Pacific Northwest National Laboratory.
▽ Computational Sciences and Mathematics Division, Pacific Northwest National Laboratory.
○ Environmental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory.
¶ Present address: University of Tennessee Medical Center, Knoxville, TN 37920.
|| Physical Sciences Directorate, Oak Ridge National Laboratory.

10.1021/pr8001832 CCC: $40.75 © 2008 American Chemical Society

Published on Web 07/01/2008

elucidate larger sets of interactions that involve a large fraction of the proteome, generally with higher speed, but less rigor, than classical methods. These higher-throughput techniques include the yeast two-hybrid technique,[14–17] arrays of immobilized proteins,[18,19] and phage display.[20] Mass spectrometry-based measurements of protein−protein interactions across the proteome in *Saccharomyces cerevisiae*[21,22] and in *Escherichia coli*[23,24] have been reported. Desirable attributes of the latter approach include direct detection of proteins (as opposed to mRNA, for example), additional experimental tools for reducing false-positive interactions,[25] and the potential to study post-translational modifications.

It would be desirable to extend application of a mass spectrometry-based approach for discovery of protein−protein interactions to a wider variety of environmentally relevant Gram-negative bacterial species, expecially those of potential interest for energy production, bioremediation, or carbon sequestration.[26] To address this need, we have developed a straightforward approach that is general to Gram-negative bacteria, and flexible enough to accommodate different affinity tags without the need to amplify and reclone all genes of interest from genomic DNA. Our goal is a system for efficient *in vivo* expression of a range of affinity tagged fusion proteins across a variety of species, particularly in those strains in which chromosomal integration is difficult or impossible. This report describes the construction and testing of a vector based on pBBR1MCS5, which has been demonstrated to replicate in a number of Gram-negative bacteria.[27] We have modified this vector to incorporate the Gateway cloning system,[28] which allows facile introduction of a library of cloned genes (entry vectors) into constructs encoding several different affinity (or other) tags (destination vectors). Using the RNA polymerase complex as a test case, we evaluated the performance of the plasmid-encoded hexahistidine/V5 epitope fusion proteins in *Rhodopseudomonas palustris*,[29] *Shewanella oneidensis* MR-1,[30] and *E. coli*,[31] and compared this tagging and purification strategy with the chromosome-encoded tandem affinity purification (TAP) tag[32] in *E. coli*.

Materials and Methods

Strains, Plasmids, and Media. Bacterial strains and plasmids used in this study are listed in Table 1. *R. palustris* CGA010[33] and pBBR1MCS5 broad-host-range vector[27] were gifts from Caroline Harwood, University of Washington-Seattle. *S. oneidensis* MR1 was obtained from ATCC (Manassas, VA). The *E. coli* strains expressing proteins from the tandem affinity purification (TAP) vector[32] integrated into the *E. coli* chromosome were a gift from Andrew Emili, University of Toronto. *E. coli* and *S. oneidensis* strains were cultured on Luria−Bertani (LB) broth or LB agar plates for cloning and strain maintenance. For affinity isolation experiments, *E. coli* was cultured on M9 media and *S. oneidensis* was grown in a defined mineral medium.[34] *R. palustris* was grown anaerobically in the light at 30 °C in a defined mineral medium containing 10 mM succinate (PMS-10).[35] Plasmid-containing strains were routinely grown with antibiotics at the following concentrations: ampicillin, 100 μg/mL; chloramphenicol, 30 μg/mL; and gentamycin, 10 μg/mL for *E. coli* and *S. oneidensis* strains, and 100 μg/mL for *R. palustris* strains. The growth medium for *E. coli* B2155 was supplemented with 100 μg/mL diaminopimelic acid (DAP).

Construction of pBBR5DEST42 Broad-Host-Range Destination Vector. Isolated pBBR1MCS5 plasmid DNA was digested with KpnI and PvuI, phenol-extracted, and treated with calf intestine alkaline phosphatase, and a 4.5 kb fragment was purified on an agarose gel. The recombination region on pET-DEST42 vector DNA (Invitrogen, Carlsbad, CA) was PCR-amplified using the following primers that include KpnI and PvuI restriction sites, respectively: 42F, 5′-cag cgg tac ctc tcg atc ccg cga aat taa ta-3′; and 42R, 5′-ctg tcg atc gct gta ggc ata ggc ttg gtt atg-3′. The PCR reactions were done in 100 μL volumes with PfuTurbo DNA polymerase (Stratagene, La Jolla, CA) and 50 ng of pET-DEST42 DNA. PCR products were analyzed on an agarose gel to verify that the products were the correct size. The PCR product was digested with KpnI and PvuI and agarose gel-purified. A total of 60 ng of PCR product was ligated with 30 ng of 4.5 kb pBBR1MCS5 KpnI/PvuI fragment following standard protocols.[36] Ligation products were transformed into *E. coli* DB3.1 chemically competent cells (Invitrogen) and transformants were selected by plating on LB agar plates containing appropriate antibiotics. Individual colonies were grown overnight in LB containing 30 μg/mL chloramphenicol and 10 μg/mL gentamycin, and plasmid DNA was prepared using QIAprep spin miniprep following the manufacturer's protocol (Qiagen, Valencia, CA). Plasmid DNA was digested with KpnI and PvuI and digestion products were analyzed on an agarose gel to confirm the presence of products of the expected sizes.

Cloning of Target Genes into Gateway Entry Vector. Genes of interest were PCR-amplified from *R. palustris* CGA010, *S. oneidensis* MR1 or *E. coli* K12 MG1655 genomic DNA with Expand DNA polymerase (Roche Diagnostics, Indianapolis, IN), and appropriate primers (Table 1). The primers contain appropriate *attB* recombination sequences, Shine−Delgarno sequence, and the ATG start codon; the transcriptional stop codon was removed to allow C-terminal fusion of affinity tag sequences. PCR products were analyzed on an agarose gel and purified with a QIAquick PCR purification kit (Qiagen). PCR product was recombined with pDONR221 DNA using BP Clonase II enzyme mix following the manufacturer's protocol (Invitrogen), and then transformed into chemically competent DH5α cells and plated onto LB with appropriate antibiotic selection. The inserts were confirmed by sequencing using M13 forward and reverse primers (Laragen, Inc.; Los Angeles, CA).

Creation of Expression Clones from Gateway Entry Clones. Entry clones in the pDONR221 vector were recombined with the pBBR5DEST42 destination vector using LR Clonase II enzyme mix following the manufacturer's protocol (Invitrogen), transformed into BL21(DE3) chemically competent cells, and plated onto selective media. Individual colonies were grown overnight in LB containing gentamycin, and plasmid DNA was purified with a Qiaprep miniprep kit (Qiagen).

Plasmid Transformation of *R. palustris* and *S. oneidensis*. To obtain electrocompetent cells of *R. palustris* CGA010, 50 mL of cultures were grown anaerobically in the light at 30 °C in PMS-10 to OD of 0.4 at 660 nm. Cells were recovered by centrifugation at 10 000 rpm for 10 min at 4 °C. Three cycles of washing and resuspension in 50 mL of ice-cold sterile distilled water were performed followed by final resuspension in 1 mL of chilled 10% glycerol. Aliquots were stored at −80 °C until needed. For optimal transformation, 40 μL of competent cells, along with 1−10 ng of plasmid DNA in 80 μL of chilled distilled water, was added to 2 mm cuvettes of a Gene Pulser (Bio-Rad Laboratories, Hercules, CA). Electroporation was performed using the following settings: voltage 2.5 kV; capacitance 25 μF; resistance 100 Ω. After electroporation 1 mL PMS-10 was added and cells were incubated at 30 °C for 20 h.

Table 1. Bacterial Strains, Plasmids and Primers Used in This Study

strain, plasmid, or primer	genotype, phenotype, or sequence of primer (5′–3′)	reference, origin, or description
E. coli		
K12	K-12 MG1655 Wild-type strain	ATCC
DH5α	F^- ϕ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	Novagen
DB3.1	F^- *gyr*A462 *end*A1Δ(sr1-*rec*A) *mcr*B *mrr hsd*S20(r_B^-, m_B^-) *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(Sm^R) *xyl*-5λ- l*eu mtl*1	Invitrogen
β2155	*thrB1004 pro thi strA hsdS lacZ*D*M15* (F9 *lacZ*D*M15 lacI*q *traD36 proA*1 *proB*1) D*dapA*::*erm* (Ermr) *pir*::RP4 [::*kan* (Kmr) from SM10]	ref 37
BL21(DE3)	F^- *omp*T *hsd*S_B($r_B^-m_B^-$) *gal dcm* (DE3)	Invitrogen
R. palustris CGA010	hupV repaired derivative of CGA009	ref 33, 29
S. oneidensis MR1	Wild-type strain	ATCC 700550[54]
Plasmids		
pET-DEST42	Ap^r, Cm^r, C-terminal 6×His and V5 epitope	Invitrogen
pBBR1MCS5	Gm^r, *mob*, broad host range cloning vector	ref 27
pBBR5DEST42	Gm^r, C-terminal 6×His and V5 epitope	This study
pDONR221	Km^r, gateway entry vector Gm^r, N-terminal GST	Invitrogen
PCR Primers[a]		
RPA3226-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGACGATCCAGAAAAATTGGC	*rpoA*
RPA3226-rev	ggg gac cac ttt gta caa gaa agc tgg gtcGTAGTGATCTTCGAAGCGCTT	
RPA0367-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCCCGTGCTGCTACTCT	*rpoH*
RPA0367-rev	ggg gac cac ttt gta caa gaa agc tgg gtcGTGCGCTGCTTCCAGCGCC	
RPA3267-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCGATGAACCAGGAAATTA	*rpoC*
RPA3267-rev	ggg gac cac ttt gta caa gaa agc tgg gtcCTCTGCCGGCGGCAGCGA	
RPA3268-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCGCAGCAGACGTTCAC	*rpoB*
RPA3268-rev	ggg gac cac ttt gta caa gaa agc tgg gtcCTCGGCAGCCTCGGCCGG	
ECK3282-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat aga ATGCAGGGTTCTGTGACAG	*rpoA*
ECK3282-rev	ggg gac cac ttt gta caa gaa agc tgg gtc CTCGTCAGCGATGCTTGCC	
SO0256-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGCAGGGTTCTGTTACAGAAT	*rpoA*
SO0256-rev	ggg gac cac ttt gta caa gaa agc tgg gtcTAGGTCGTCTGCTAAACTAGCT	
SO0225-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGTTTACTCCTATTCTGAAA	*rpoC*
SO0225-rev	ggg gac cac ttt gta caa gaa agc tgg gtcTTCCTGATCCAACTCGATATTA	
SO0360-for	gggg aca agt ttg tac aaa aaa gca ggc ttc gaa gga gat agaATGGCTCGCGTAACTGTAGAAG	*rpoZ*
SO0360-rev	ggg gac cac ttt gta caa gaa agc tgg gtcTAGTGAACGGCCTTCAGCGAT	

[a] Gene specific sequences are in upper case. Shine–Delgarno sequence is underlined.

Transformants were selected by plating on PMS-10 agar plates containing 100 μg/mL gentamycin and incubated anaerobically in the light at 30 °C. For transformation of *S. oneidensis*, purified plasmids were first transformed into chemically competent *E. coli* β-2155.[37] Plasmids were then transferred from *E. coli* β-2155 to *S. oneidensis* MR-1 by conjugation and transformants were selected on LB agar plates containing appropriate antibiotic.

Western Blots. To test for protein expression, individual colonies of expression strains were grown in appropriate liquid media as described previously to midlog phase and cells were harvested by centrifugation. Cell pellets were lysed in phosphate buffered saline (PBS) with 1 μL/mL BugBuster plus Benzonase Nuclease (Novagen, Madison, WI), 10 μg/mL lysozyme, 100 μg/mL phenylmethylsulfonylfluoride (PMSF) and 10 μg/mL leupeptin. After mixing gently for 20 min at room temperature, lysates were centrifuged at 16 000*g* for 15 min at 4 °C. Cleared lysates were separated on 10% SDS-PAGE and transferred to Immobilon P membrane (Millipore, Billerica, MA) in Tris-glycine-MeOH buffer for 1 h at 200 mA. Membranes were washed overnight at 4 °C in PBS containing 0.1% Tween 20 (PBS-Tween). Nonspecific binding was blocked by incubation of the membrane with gentle rocking for 1 h in 5% nonfat milk powder, 10% bovine serum in PBS-Tween ("Blotto"). The membrane was then incubated for 2 h with rocking at room temperature with the primary antibody, mouse anti-V5 epitope (Invitrogen) at 1/5000 dilution in Blotto. After extensive washing in PBS-Tween, the membrane was incubated with rocking for 1.5 h in horseradish peroxidase-conjugated goat anti-mouse IgG (Bio-Rad, Hercules, CA) at 1/1000 dilution. After washing, color was developed by the addition of ImmunoPure metal enhanced DAB substrate (Pierce, Rockford, IL).

Affinity Isolation. Cultures of 0.8 L of each bacterial strain were grown in appropriate media to mid-logarithmic phase, cells were harvested by centrifugation, and pellets were frozen at −80 °C until needed. Frozen cell pellets were thawed, and cells were lysed using 5 vol of lysis buffer (w/v), prepared with

10× BugBuster extraction reagent (Novagen) diluted to 1× in Ni-NTA binding buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8), containing 1 μL/mL benzonase nuclease reagent (Novagen), lysozyme (5 kU/g cell paste), 100 μg/mL PMSF and 10 μg/mL leupeptin. Cells were suspended in the lysis buffer, and incubated with end-over-end rotation at room temperature for 20 min. Lysates were centrifuged for 30 min at 12 000*g* for 20 min. The supernatant was transferred to a fresh tube, centrifuged for 15 min at 20 000*g* at 4 °C, and the resulting supernatant (lysate) was used for subsequent purification steps.

The lysate was added to 200 μL of a 50% suspension of Ni-NTA agarose beads (Invitrogen), which had previously been washed 3−4 times with Ni-NTA binding buffer. The lysate and Ni-NTA beads were incubated with end-over-end tube rotation for 1 h at room temperature, and then centrifuged at 100*g* for 30 s and the supernatant removed. The beads were washed three times with 1 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 100 μg/mL PMSF, 10 μg/mL leupeptin). Proteins were eluted from the beads using four 50 μL aliquots of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, 100 μg/mL PMSF, 10 μg/mL leupeptin). The eluates were combined and 400 μL of water containing 100 μg/mL PMSF, and 10 μg/mL leupeptin was added. After a final short clearing spin to pellet any remaining beads, ∼90% of the supernatant was transferred to a clean tube.

The second purification step was performed using mouse anti-V5 monoclonal antibody agarose beads (Sigma). Eluate from the Ni-NTA purification was added to 100 μL of 50% slurry of anti-V5 beads, which had previously been washed four times with PBS. Sample and beads were incubated for 1 h with end-over-end rotation of the tube at room temperature. Three cycles of brief centrifugation at 100*g* for 30 s followed by addition of 1 mL of PBS, gentle shaking, and removal of the supernatant, were performed. Proteins were eluted from the anti-V5 beads with three 50 μL vol of aqueous 80% acetonitrile containing 1% formic acid. After centrifugation at 100*g* for 30 s to pellet any remaining beads, supernatant was transferred to a fresh tube.

For TAP affinity isolations, *E. coli* cellular lysates were incubated with IgG (Amersham/GE) affinity resin for 1 h. Elution from the IgG resin was performed by treatment with AcTEV protease (Invitrogen) for 1 h. A second 1 h incubation with calmodulin affinity resin (Amersham/GE) further isolated the affinity-tagged proteins. Proteins were eluted from the calmodulin affinity resin as described by Zeghouf et al.[38]

Trypsin Digestion. Proteins eluted from affinity resins were dried by vacuum centrifugation, then reconstituted in 150 μL of 50 mM Tris, 10 mM $CaCl_2$ buffer (pH 7.6). One microgram of trypsin (sequencing grade; Promega) was added and the digestion was allowed to proceed overnight at 37 °C. The sample was acidified with 50 μL of 2.5% trifluoroacetic acid (TFA). Two 65 μL volumes were aliquoted into separate tubes for desalting via solid-phase extraction (OMIX C18, 100 μg size, Varian) using the manufacturer's protocol. Peptides were eluted from the OMIX tip in 100 μL aqueous 50% acetonitrile, 0.1% TFA. A 100 μL portion of aqueous 0.1% TFA was added, and the sample volume was decreased by approximately one half using vacuum centrifugation. Particulates were removed by centrifugation for 15 min at 10 600*g* before transfer to autosampler vials.

Mass Spectrometry and Protein Identifications. The resulting peptides were analyzed by automated LC-ESI-MS/MS as reported previously in greater detail.[39] Briefly, an LCPackings (Sunnyvale CA) chromatographic system (Famos autosampler, Switchos flow switching module, and Ultimate HPLC) was interfaced with a ThermoFinnigan (San Jose CA) DecaXP Plus quadrupole ion trap mass spectrometer via a nanospray source. After injection via the autosampler, peptides were concentrated onto a reverse-phase preconcentration column and washed to remove excess salt. The peptides were then eluted onto a 75 μm i.d. C18 reverse-phase analytical column. Peptides were eluted from the analytical column using a gradient from 95% water/5% acetonitrile to 30% water/70% acetonitrile (with 0.1% formic acid in all mobile phases). Blank chromatographic runs between samples minimized carryover. The mass spectrometer was operated in a data-dependent MS/MS mode, with up to four peptides automatically selected per MS scan for MS/MS analysis. Proteins interacting with the bait proteins were identified from MS/MS spectra of their tryptic peptides via Sequest.[40] The resulting peptide identifications and other relevant statistics were filtered and collated by protein using DTASelect.[41] Peptide identifications required standard minimum XCorr values of ≥1.8 for singly charged, ≥2.5 for doubly charged, or ≥3.5 for triply charged ions, and DeltaCN ≥0.08. Proteins were initially identified based on detection of one or more peptides and subsequently filtered at later stages of data analysis. Sequest searches were performed against separate FASTA protein databases for the three bacterial species, compiled as follows. Sequences for 4833 *R. palustris* proteins were obtained from the ORNL microbial genome annotation pipeline at http://www.ornl.gov/sci/GenomestoLife/index.shtml,[29] with an additional 44 common contaminants and standard proteins. The *E. coli* protein sequence file contained 4337 proteins,[42] plus their sequence-reversed counterparts and 188 common contaminants and standard proteins. The *S. oneidensis* MR-1 file contained 4898 *S. oneidensis* protein sequences,[43] plus 44 common contaminants and standard proteins.

Evaluation of Protein−Protein Interactions. The BEPro3 algorithm[44] was applied for statistical evalution of observed bait−prey interactions. Two BePro3 results were used: the *Bayes Odds*, a score associated with each bait−prey pair, ranging from 0 (low likelihood of interaction) to 1 (high likelihood of interaction); and the *Ubiquity*, a measure of the tendency of a given prey protein to be detected in association with a large variety of baits (0 indicates a "selective" prey, and 1 indicates a "sticky" prey.) For the *R. palustris* proteins, the data set included 77 LC-MS/MS measurements using 18 different affinity-tagged baits.[44] The *S. oneidensis* data set encompassed 84 LC-MS/MS measurements of affinity isolations obtained using 29 different affinity-tagged baits. BEPro3 calculations were not performed for *E. coli* results due to the small size of the data set.

Results

Figure 1 outlines the approach tested in this paper for systematically identifying interactions among proteins across a range of bacterial species. The major steps are PCR amplification of genes of interest, recombination of each gene into a Gateway entry vector, recombination of each entry clone into a pBBR5DEST42 destination vector, transformation of host cells, cell culture and harvest, affinity isolation, mass spectrometric protein identification, and data analysis.

To demonstrate our system, we focused on the protein complex RNA polymerase, which has been suggested as a

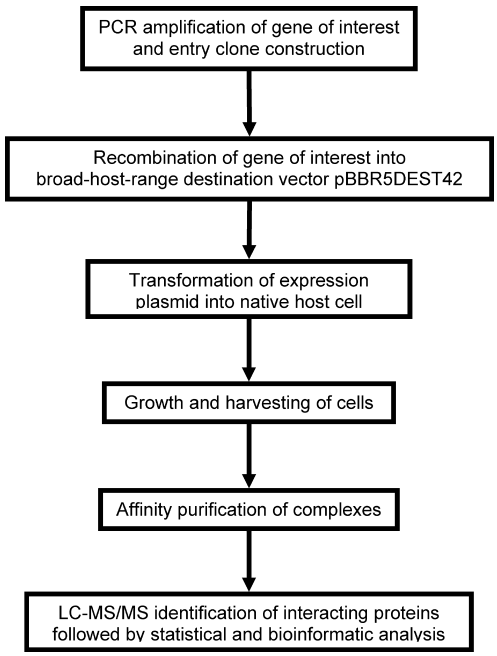

Figure 1. Overview of our approach for experimental determination of protein–protein interactions in Gram-negative bacteria.

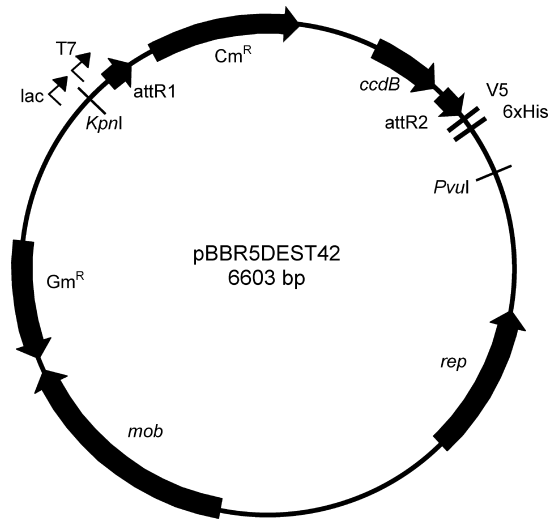

Figure 2. Map containing relevant features of pBBR5DEST42 destination vector (see Materials and Methods for details of plasmid construction) *rep*, pBBR1MCS5 replication protein; *mob*, mobilization protein; *ccdB*, death gene; CmR, chloramphenicol resistance protein; GmR, gentamicin resistance protein; V5, C-terminal V5 epitope fusion; 6×His, hexahistidine C-terminal fusion.

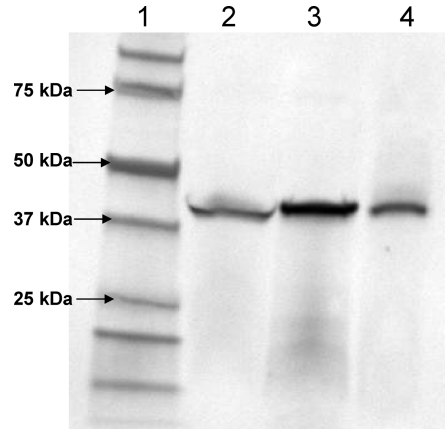

Figure 3. Western blot analysis of several RpoA hexahistidine/V5-fusion proteins expressed from pBBR5DEST42. Lane 1, MW standards; lane 2, *E. coli* K12 pBBR5DEST42-ECK3282, predicted MW 40 768 Da; lane 3, *R. palustris* pBBR5DEST42-RPA3226, predicted MW 41 791 Da; lane 4, *S. oneidensis* MR1 pBBR5DEST42-SO0256, predicted MW 40 417 Da. Each lane contains ~30 μg of total cell lysate. The blot was probed with anti-V5 antibody.

standard for evaluating methods for measuring protein–protein interactions.[45] DNA-directed RNA polymerase uses DNA as the template for RNA synthesis during transcription. The "core" bacterial RNA polymerase contains two copies of the α subunit (RpoA) along with β (RpoB), β' (RpoC), and ω (RpoZ in *E. coli*; RnpO in *R. palustris*) subunits. This core can perform transcription of DNA into RNA, but has little specificity. Incorporation of additional protomers, the sigma factors, which join the complex depending on the cell's stress or environment,[46] provides specificity in recognition of promoter regions on DNA. As described below, detection of the core components, sigma factors and other proteins known to interact with RNA polymerase provided a basis for evaluating our approach.

Vectors Coding for Affinity-Tagged Proteins. Figure 2 shows the plasmid map for the pBBR5DEST42 destination vector, created by insertion of pET-DEST42 recombination region into the vector backbone of pBBR1MCS5. This new vector combines the convenience of facile introduction of genes of interest using the Gateway site-specific recombinase system with the medium copy number and broad host range conferred by the pBBR1MCS5 origin of replication. The plasmid also encodes proteins necessary for mobilization of the plasmid by RK2 conjugal transfer from an appropriate host strain, the *lac* and T7 promoter sequences for expression in *E. coli* or other appropriate host strains, and a carboxy-terminal hexahistidine/V5 epitope fusion for affinity isolation. To evaluate the ability of this vector to stably replicate and drive expression, we introduced destination plasmids bearing the *rpoA* gene from three phylogenetically diverse bacterial species, the alpha proteobacterium *R. palustris*, the gamma-proteobacteria *S. oneidensis* and *E. coli*, and subsequently transformed the respective vectors into native host strains. Expression of plasmid-encoded fusion proteins was confirmed by Western blot analysis. Figure 3 demonstrates expression of hexahistidine/V5-tagged RpoA from *E. coli* containing pBBR5DEST42-ECK3282 (lane 2), from *R. palustris* cells containing pBBR5DEST42-RPA3226 (lane3) and from *S. oneidensis* cells containing pBBR5DEST42-SO0256 (lane 4).

Comparison of Chromosomal and Plasmid-Encoded SPA Affinity Tags in *E. coli*. Table 2 compares results from several different affinity isolation strategies using the α subunit, RpoA, of RNA polymerase as the tagged protein. The data presented include results from large-scale protein–protein interaction studies in *E. coli* using the hexahistidine tag,[23] or a dual affinity tag (SPA or TAP),[24] as well as results obtained in the present study using the TAP-tagged chromosomal insert[24] and the plasmid-encoded hexahistidine/V5-tagged proteins as baits. Results from the present study are presented in Table 2 as the number of MS/MS spectra assigned to a particular protein (the "spectrum count"), which is a crude indicator of the amount of that protein in the sample.[47] For comparison, Table 2 also

Table 2. Proteins Co-Isolating with Affinity-Tagged RpoA in *E. coli*[a]

Tag, Location	6-His/V5,[b,c]plasmid	TAP[b,d], chromosomal	TAP/SPA, chromosomal	6-His, plasmid		
Detection Method	LC-MS/MS	LC-MS/MS	LC-MS/MS and MALDI	MALDI		
source	this study	this study	ref 24[e]	ref 23[f]	STRING[g]	Prolinks[h]
Protein[I]						
RpoC	62.5	381	**Y**	y	0.999	0.505 GN
RpoB	50.5	331	**Y**	y	0.999	0.53 GN
RpoD	6	62	**Y**	y	0.999	0.577 PP
RplB	15.5	6	**Y**	y	0.909	0.807 GN
RpsE	30	5	**Y**	y	0.961	0.877 GN, 0.17 PP
RpsD	5.5	4	**Y**	y	0.989	0.87 PP, 0.853 GN, 0.676 GC
RpoA (bait)	107	92	y			
RpoZ	11	19	**Y**		0.998	N
RplD	3.5	5	**Y**		0.926	0.799 GN, 0.392 PP
RpsG	4	3	**Y**		0.898	0.717 GN, 0.354 PP
RpsJ	5	2	y		0.923	0.795 GN, 0.471 PP
RpsM	4	2	y		0.98	0.879 GN, 0.637 PP
RplO	6		y	y	0.965	0.745 GN
TufA	24.5	23			0.85	0.686 GN
HupA		4	y		0.16	0.167 PP
NusA		4	**Y**		0.994	0.778 PP
RpsO		2	y		N	0.248 PP
RpsT		2	y		N	N
HepA			**Y**	y	0.579	N
Rho	8		y		0.535	0.274 PP
RplA	5		y		0.871	0.689 PP, 0.367 GN
RplE	18.5		y		0.977	0.846 GN, 0.195 PP
RplF	2			y	0.948	0.867 GN, 0.633 PP
RplM	6.5		**Y**		0.86	0.685 GN, 0.388 PP
RplP	6.5		y		0.947	0.821 GN, 0.633 PP
RpsB	26.5		**Y**		0.687	0.435 PP
RpsC	17.5		**Y**		0.974	0.816 GN
GreB			**Y**		0.897	0.148 PP
NusG			**Y**		0.899	0.566 PP, 0.454 GN
RplC			**Y**		0.899	0.798 GN, 0.314 PP
RplL			**Y**		0.919	0.159 GN
RpsP			**Y**		0.242	N

[a] Table contains proteins detected by two or more approaches, plus those reported as validated by Butland et al.[24] [b] Table entries correspond to number of MS/MS spectra identified as arising from peptides of the corresponding protein. Average of two LC-MS/MS runs for His-V5 tag; single run for TAP tag. [c] Proteins detected only with His-V5 tagged RpoA: AtpF (6), Crp (5.5), CyoA (7.5), DacB (3), DnaJ (10.5), GatY (9), GlpD (36), LldD (3), PpiD (4), RplJ (8), RpsI (4.5), RpsK (4), SecG (3), XerD (6), YfgM (2.5), YqjD (8.5), YqjI (4.5). This list provides gene name followed in parentheses by number of MS/MS spectra (see note b above). [d] Proteins detected only with TAP tagged rpoA in present study: Dps (23), HupA (4), HupB (4), PepA (12), RplV (2), RpmC (6), RpoE (2), RpoS (10), SeqA (4), RpsO (2), RpsT (2). This list provides gene name followed in parentheses by number of MS/MS spectra (see note b above). [e] Interactions reported by Butland et al.[24] denoted by "y", with "**Y**" denoting interactions reported as validated. Other interactions reported:[24] AceE, AceF, AspS, CspC, MreB, PssA, RfaK, RplS, RplU, RplW, RplX, RpmA, RpmB, RpmG, RpsA, RpsF, RpsN, RpsS, RpsU, WecG. [f] Interactions reported by Arifuzzaman et al.[23]denoted by "y" in the table. Other interactors reported:[23] Fur, GreA, HybE, Pta, YccC. [g] Combined score predicted from STRING (http://string.embl.de). N: not found. [h] Predicted scores from Prolinks (http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav), with prediction methods: PP, phylogenetic profile; GN, gene neighbor; GC, gene cluster. N, not found.[49] [I] Bold entries are components of RNA polymerase core enzyme.

includes predictions from STRING[48] (http://string.embl.de), and from Prolinks (http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav).[49] The data resulting from the four different experimental approaches shown in Table 2 detected RNA polymerase core components β and β' (RpoB and RpoC), as well as the most common sigma factor, σ^{70} (RpoD). Results from both tags used in the present study, as well as literature results obtained using the TAP or SPA tag,[24] include the ω subunit (RpoZ); RpoZ was not reported using the hexahistidine tag alone.[23] With the use of the TAP or SPA tag, previous studies reported detection of both NusA and NusG,[24] which are known to interact with RNA polymerase.[50] Using our detection system with the chromosomally encoded TAP-tagged strain,[24] we detected NusA but not NusG. Neither NusA nor NusG was detected using the hexahistidine tag alone[23] or the hexahistidine/V5 tag (this study). Among other proteins in Table 2 known to associate with RNA polymerase, and detected in the published studies[23,24] (HepA, GreB,[51] and Rho[52]), we detected Rho with the hexahistidine/V5 tag, but not with the TAP tag. All four approaches led to detection of ribosomal proteins RplB, RpsE, and RpsD; several other ribosomal proteins were detected using one or more approaches. All four approaches also led to detection of several proteins that are not obviously related to RNA polymerase (see also footnotes b and c, Table 2).

RNA Polymerase in *S. oneidensis*. Table 3 summarizes protein−protein interactions determined using RpoA, RpoC and RpoZ expressed as affinity-tagged fusions from the pBBR5DEST42 plasmid in *S. oneidensis*. A prey protein is listed in Table 3 only if it fulfills all of the following conditions for at least one bait: BePro3 *Ubiquity* ≤ 0.5, BePro3 *Bayes Odds* ≥ 0.5, with two or more peptides identified. A more complete listing is provided in Supplementary Table 1. Bold entries in Table 3 identify bait−prey interactions with BePro3 *Bayes Odds* ≥ 0.5. Multiple bold entries among the core components of the RNA polymerase enzyme indicate the specificity of the measurement. No sigma factors were detected with these baits for *S. oneidensis*. The RpoA bait showed the most interactors, including several that are not known components of RNA polymerase, but are abundant proteins in *E. coli* (ribosomal proteins, β subunit of ATP synthase F1). It is noteworthy that the average spectrum count for the RpoA bait protein itself was, by a large margin, the highest among the baits in Table 3, suggesting strong preferential overexpression of this protein from the plasmid. Unlike the other affinity-tagged components

Table 3. Selected[a] Proteins Detected in Affinity Isolations with Tagged Components of RNA Polymerase in *S. oneidensis* MR-1

Preys				Baits					
				SO0225, RpoC		SO0256, RpoA		SO0360, RpoZ	
				3 BR, 6 TR[b]		4 BR, 6 TR[b]		3 BR, 6 TR[b]	
gene ID	name[c]	product	*Ubiquity*	*BO*	SC_{av}	*BO*	SC_{av}	*BO*	SC_{av}
SO0224	rpoB	DNA-directed RNA polymerase, beta subunit	0.17	**1**	**23.33**	**1**	**104**	**1**	**10.83**
SO0225	rpoC	DNA-directed RNA polymerase, beta subunit	0.21	**1**	**16.33**	**1**	**81**	**1**	**6.67**
SO0256	rpoA	DNA-directed RNA polymerase, alpha subunit	0.17	**1**	**17.17**	**1**	**42.67**	**1**	**7.83**
SO0360	rpoZ	DNA-directed RNA polymerase, omega subunit	0.09	**0.59**	**1.33**	**1**	**12**	**1**	**17.33**
SO1490	adhB	alcohol dehydrogenase II	0.28	**0.98**	**2**	0.25	**8**	**0.96**	**1.83**
SO4236	leuA	2-isopropylmalate synthase	0.35			**1**	**4.83**	**1**	**3.67**
SO3907		conserved hypothetical protein	0.04	0	1.17	0	1.17	**0.99**	2
SO4747	atpD	ATP synthase F1, beta subunit	0.24			**0.96**	**3.5**		
SO1624	purU	formyltetrahydrofolate deformylase	0.48			**1**	**2.33**	**1**	**1**
SO1630	tsf	translation elongation factor Ts	0.48			**1**	**2.83**		
SO4263		conserved hypothetical protein	0.24			**1**	**2.83**		
SO0227	rpsG	ribosomal protein S7	0.24			**1**	**2.33**		
SO0253	rpsM	ribosomal protein S13	0.37			**1**	**2.17**		
SO0222	rplJ	ribosomal protein L10	0.38			**1**	**2**		
SO2485		deoxyguanosinetriphosphate triphosphohydrolase, putative	0.29			**0.94**	**2**		
SO3652	rplU	ribosomal protein L21	0.36			**0.66**	**1.83**		
SO2299	thrS	threonyl-tRNA synthetase	0.19			**0.87**	**1.67**		
SO3348	hemH-2	ferrochelatase	0.12			**0.93**	**1.33**		
SO4105	mshA	MSHA pilin protein MshA	0.37			**0.98**	**1.17**		

[a] Selection criteria: *Ubiquity* ≤ 0.5, *Bayes Odds* 0.5, two or more peptides. See Supplementary Table I for complete listing. [b] BR: number of different cultures (Biological Replicates); TR: number of LC-MS/MS experiments (Technical Replicates). *BO*, *Bayes Odds*; SC_{av}, spectrum count averaged across technical replicates Bold entries correspond to *Bayes Odds* ≥ 0.5. [c] Underlined entries are components of RNA polymerase core enzyme.

of RNA polymerase that were used as bait in *S. oneidensis*, RpoB was not detected when used as bait, based on duplicate LC-MS/MS measurements from a single culture. However, this protein was detected as a prey in experiments where RpoA, RpoC, or RpoZ was the affinity-tagged bait. While Table 3 lists the more obvious interactors with the RNA polymerase core components, Supplementary Table I contains evidence for additional interactions. For example, prey proteins that were observed with the RpoA bait, but did not meet all the criteria for inclusion in Table 3, include NusA and NusG, which are known interactors of RNA polymerase.[50]

RNA Polymerase in *R. palustris*. Table 4 summarizes protein–protein interactions for DNA-dependent RNA polymerase in *R. palustris*, using α (RpoA, RPA3226), β (RpoB, RPA3268), and β' (RpoC, RPA3267) subunits of the core RNA polymerase enzyme and the heat shock sigma factor σ^{32} (RpoH, RPA0367) expressed as affinity-tagged fusions from the pBBR5DEST42 plasmid. A prey protein was listed in Table 4 only if it fulfilled the conditions described above for Table 3; a more complete listing is provided in Supplementary Table II. Bold entries in the table identify bait–prey interactions with BePro3 *Bayes Odds* ≥ 0.5. As noted above for *S. oneidensis*, multiple bold entries suggest robust detection of interactions among the core components of the enzyme. Although the ω subunit of the core complex from this organism, RnpO, was not successfully cloned and expressed as an affinity-tagged fusion, this protein was nonetheless detected in experiments using the other RNA polymerase components as baits. The major sigma factor (σ^{70}, RpoD) was detected using several baits. Using σ^{32} as affinity-tagged bait led to detection of several core components. Also interacting with σ^{32} were heat shock proteins GrpE (RPA0331, Table 4) and, less strongly, DnaK (RPA0333, Supplementary Table II); these two proteins have been reported to associate with σ^{32} in *E. coli*.[53] A 14.5 kDa acidic conserved protein of unknown function, RPA0060, was detected fairly robustly with the RpoA bait, suggesting a previously unreported association with the RNA polymerase complex. RPA0060 was also observed, albeit weakly, as a prey for both the RpoC and RpoB baits. Several abundant proteins not necessarily associated with RNA polymerase were observed, including ribosomal proteins, aconitase, and ATP synthase subunits.

Discussion

Microbes inhabit a diverse array of habitats, deriving energy from a variety of sources and biochemicals from a wide range of raw materials. The protein complexes and biochemical pathways that microbes have evolved over countless iterations in nature represent an amazing range of activities and strategies for bioremediation and energy production.[26] To gain access to this potentially vast resource, it will be necessary first to develop faster and more accurate methods, such as that described in this report, for studying the protein–protein interactions involved in these complexes and pathways.

To enhance our ability to discover such novel and potentially useful biochemical capabilities in microbes, we have developed a strategy to extend studies of protein–protein interactions to a larger subset of bacterial species. The basis of this strategy is the pBBR5DEST42 vector, which can be introduced into a broad range of Gram-negative bacterial hosts (Figure 2). In addition to broad species applicability, the strategy facilitates expression of a variety of fusion protein types. Because the approach is based on the Gateway system, each protein of interest can be "decorated" with new affinity or other types of tags simply by constructing a new destination vector; each gene of interest need only be cloned and verified once. This system is particularly useful in species where transformation is feasible, but chromosomal integration is not. We demonstrated that pBBR5DEST42 could be used to express fusion proteins in *R. palustris*, *S. oneidensis* and *E. coli* (Figure 3, Tables 2–4).

Initial large-scale studies of protein–protein interactions in yeast used different approaches for producing affinity-tagged protein. In one study, engineered genes encoding each protein of interest and an affinity tag were introduced into a chromosome by homologous recombination,[21] while another group

Table 4. Selected[a] Proteins Detected in Affinity Isolations with Tagged Components of RNA Polymerase in *R. palustris*

Preys				Baits							
				RPA0367, RpoH		RPA3226, RpoA		RPA3267, RpoC		RPA3268, RpoB	
				3 BR, 6 TR[b]		4 BR, 4 TR[b]		2 BR, 4 TR[b]		4 BR, 10 TR[b]	
gene ID	name[c]	product	*Ubiquity*	*BO*	SC_{av}	*BO*	SC_{av}	*BO*	SC_{av}	*BO*	SC_{av}
RPA3268	rpoB	RNA polymerase beta subunit	0.19	0.02	19.83	**0.99**	**177**	**0.99**	**130.25**	**1**	**85.1**
RPA3267	rpoC	RNA polymerase beta' subunit	0.31	**1**	**12.17**	**1**	**87.25**	**1**	**137.75**	**1**	**43.2**
RPA3226	rpoA	DNA-directed RNA polymerase alpha subunit	0.25	**1**	**8.67**	**1**	**123.5**	**1**	**43.25**	**1**	**31.8**
RPA0367	rpoH	stress response sigma factor	0.19	**1**	**38**	**0.98**	**9.5**	**0.98**	**11**	0	2.9
RPA1288	rpoD	RNA polymerase sigma subunit	0.19			**1**	**11.5**	**1**	**25.5**	**1**	**12.9**
RPA2692	rnpO	RNA polymerase omega subunit	0.19	0.01	1.33	**1**	**13.5**	**1**	**17.5**	**1**	**11.7**
RPA0183	RPA0183	unknown protein	0.19	**0.81**	**9.67**	0	2.75	0.33	15.	**0.95**	**7.5**
RPA1093	ribB	possible GTP cyclohydrolase II, riboflavin	0.07	**0.87**	**11.33**	0	2.75	0.41	13.5	0.02	4.7
RPA0333	dnaK	heat shock protein DnaK (70)	0.06	**1**	**23.5**						
RPA3254	rpsG	30S ribosomal protein S7	0.45	**1**	**5.17**	**0.81**	**2.75**	**0.99**	**10.25**	**0.92**	**2.7**
RPA3246	rpsS	30S ribosomal protein S19	0.38	**1**	**6.33**	0.01	1	0.13	3.75	**1**	**2.6**
RPA1548	puhA	H subunit of photosynthetic reaction center	0.05	0.24	5			**0.53**	**6**	0	2.5
RPA0304	RPA0304	possible outer membrane lipoprotein GNA33	0.49	**1**	**6.33**	0.01	1	0.23	1.25	0.05	1.3
RPA0060	RPA0060	conserved unknown protein	0.06			**0.9**	**7.25**	0.14	2	0	0.6
RPA1597	purQ	phosphoribosylformylglycinamide synthetase	0.12	0	0.67			**0.95**	**8.75**		
RPA4356	ctc, rplY, L25	putative 50S ribosomal protein L25	0.12	0	2.5	0	0.75	**0.87**	**5.25**		
RPA0202	acnA	aconitate hydratase	0.06	0	0.5			**0.97**	7.5		
RPA1653	RPA1653	conserved unknown protein	0.07	**0.9**	**3.83**	0.01	2.75			0	1.3
RPA3252	tufA, EF-Tu	elongation factor Tu	0.06					**0.93**	**6.5**	0	1.1
RPA3283	tuf, EF-Tu	elongation factor Tu	0.06					**0.93**	**6.5**	0	1.1
RPA3270	rplJ	50S ribosomal protein L10	0.17	**0.62**	**2.5**	0	0.5	**0.73**	**3.5**	0	0.8
RPA1205	RPA1205	putative alcohol dehydrogenase	0.12					**0.99**	**5.75**	0	1.2
RPA0844	atpF2	putative FoF1 ATP synthase, subunit B'	0.24	0	1.67	0	0.5	**0.7**	**4.75**		
RPA3238	rplE	50S ribosomal protein L5	0.06	0	0.67			**0.98**	**5.25**		
RPA0176	atpD	putative H+-transporting ATP synthase beta	0.31	0	1.5			**0.98**	**4**	0	0.4
RPA0241	rplS, L19	50s ribosomal protein L19	0.07	0	0.67			**0.99**	**4.75**		
RPA0240	leuC	3-isopropylmalate dehydratase	0.06	0	1.5			**0.95**	3.5		
RPA0331	RPA0331	possible heat shock protein (HSP-70 cofactor)	0.06	**1**	**5**						
RPA1051	RPA1051	pyruvate phosphate dikinase	0.06			0.07	0.75	**0.77**	**4**		
RPA2922	rpsB	30S ribosomal protein S2	0.13	0	1.17			**0.96**	**2.75**	0	0.5
RPA0217	sdhA	succinate dehydrogenase flavoprotein subunit	0.1					**0.63**	**2.75**		
RPA0934	RPA0934	conserved unknown protein	0.06					**0.68**	**2.5**		
RPA0854	hemO	5-aminolevulinic acid synthase (ALAS)	0.06					**0.61**	**2**		
RPA3253	fusA, EF-G	elongation factor G	0.06					**0.96**	**1.75**		
RPA4636	RPA4636	FeoA family	0.02	**0.55**	**1.67**						

[a] Selection criteria: *Ubiquity* ≤ 0.5, *Bayes Odds* ≥ 0.5, two or more peptides. See Supplementary Table II for complete listing. [b] BR: number of different cultures (Biological Replicates); TR: number of LC-MS/MS experiments (Technical Replicates). *BO*, *Bayes Odds*; SC_{av}, spectrum count averaged across technical replicates. Bold entries correspond to *Bayes Odds* ≥ 0.5. [c] Underlined entries are components of RNA polymerase core enzyme.

employed heterologous expression, with genes for affinity-tagged proteins introduced on plasmids.[22] It has been suggested that the former approach may result in fewer undetected interactions because protein expression is under the control of native promoters and at chromosomal copy number.[45] For this reason, we compared overall results from plasmid encoded versus chromosomally encoded affinity-tagged RpoA expressed in *E. coli*. To evaluate the performance of the pBBR5DEST42 vector in the context of a strategy that also includes affinity isolation and mass spectrometric identification of interacting proteins, we compared results in *E. coli* from the pBBR5DEST42-encoded fusion of RpoA with the dual hexahistidine/V5 tag to the chromosomally encoded TAP-RpoA,[24] and a plasmid-encoded hexahistidine tagged RpoA.[23] Both "dual-tag" strategies led to identification of all four components of the core enzyme, the major sigma factor, and Rho; the hexahistidine/V5 tag did not lead to detection of NusA, NusG, HepA, or GreB. The chromosomal TAP/SPA tag thus offers more sensitivity in some cases than the plasmid-borne dual hexahistidine/V5 tag. From our results, it is not possible to determine whether this is due to the location of the tagged gene (chromosomal versus plasmid) or to the properties of the tag and the associated affinity isolation. However, results shown in Table 2 suggest that, in our hands, the overall methodology, when applied to a chromosomally tagged RpoA, provides similar results for proteins expected to be more abundant in the sample while diverging somewhat for polypeptides that associate conditionally with RNA polymerase and for background proteins. Nonetheless, the plasmid-encoded hexahistidine/V5 tagging strategy used in the current study identifies known interaction partners for RpoA. The results are generally comparable (given the recognized noisy nature of these types of measurements[13]) to previous studies reported in the literature[23,24] for major interacting proteins, but with some differences in less-abundant or transient interactors and in the background proteins observed. Affinity-tagged baits expressed from plasmid-encoded genes could lead to false interactions that would not occur in the native complex due to expression of the bait protein at inappropriate times, concentrations or locations in the cell.

When we extended the plasmid-encoded hexahistidine/V5 strategy to other components of RNA polymerase in *R. palustris* and *S. oneidensis* MR-1, the expected small "networks" of multiple connections among subunits emerged. As shown in Tables 3 and 4, interactions among the core components of RNA polymerase could be confidently identified in *R. palustris* and *S. oneidensis*. Other interactors besides the core components were also detected. However, some known interactors (sigma factors for *S. oneidensis*; NusA and NusG for *R. palustris*) were not detected. The observed differences between the two species suggest that a comparative analysis may be fruitful. For example, the pBBR5DEST42 vector may cause different species-dependent stresses that affect expression and/or assembly of RNA polymerase holoenzyme. The biophysical characteristics (e.g., binding constants) of protomers that conditionally associate with the RNA polymerase core may differ between species. Regulation of RNA polymerase via sigma factors and

other components may differ between species. Examination of this complex in additional species may help to focus such a comparative analysis. As with all affinity-MS approaches, a number of proteins were detected that are not expected interactors with RNA polymerase (potential false positives), especially abundant proteins such as components of the ribosome. The BePro3 tool[44] proved useful for distinguishing potentially authentic bait–prey interactions from "background."

Conclusions

We have described a tool for expressing fusion proteins containing affinity tags via a system that combines a broad-host-range capability[27] with the flexibility to incorporate any of several different affinity tags via the commercial Gateway site-specific recombination system.[28] Although this paper has focused on the C-terminal hexahistidine/V5 affinity tag, other constructs can be readily generated; for example, we have evaluated N-terminal glutathione-*S*-transferase (GST), C-terminal GST, and N-terminal hexahistidine.

Our goal is to perform large-scale experimental studies of protein–protein interactions in microbial species that have potential for environmental applications. We are currently applying this system to large-scale characterization of protein–protein interactions in *R. palustris*. The system we describe will allow us to pursue several refinements to optimize throughput and accuracy in identifying protein–protein interactions. The flexibility of the destination vector design will allow us to explore the use of other affinity tag combinations and promoters, potentially expanding the range of protein–protein interactions that can be characterized through the powerful combination of affinity isolation and mass spectrometry.

Acknowledgment. We thank the laboratories of Caroline Harwood (University of Washington) and Andrew Emili (University of Toronto) for generous gifts of materials. This research was sponsored by the Genomics:GTL program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract no. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC. W.J.H. acknowledges the UT-ORNL Graduate School of Genome Science and Technology for support.

Supporting Information Available: Supplementary Table I and Table II, all proteins detected in affinity isolations with tagged components of RNA polymerase in *S. oneidensis* and *R. palustris*. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Alberts, B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **1998**, *92* (3), 291–294.
(2) Klotz, I. M.; Darnall, D. W.; Langerman, N. R.; Neurath, H.; Hill, R. L.; Boeder, C. L. Quaternary structure of proteins. In *The Proteins*; Academic Press: New York, 1975; Vol. 3, pp 293–411.
(3) Phizicky, E. M.; Fields, S. Protein-protein interactions—methods for detection and analysis. *Microbiol. Rev.* **1995**, *59* (1), 94–123.
(4) Alberts, B.; Miakelye, R. Unscrambling the puzzle of biological machines—the importance of the details. *Cell* **1992**, *68* (3), 415–420.
(5) Oliver, S. Proteomics: Guilt-by-association goes global. *Nature* **2000**, *403* (6770), 601–603.
(6) de Silva, E.; Stumpf, M. P. H. Complex networks and simple models in biology. *J. R. Soc. Interface* **2005**, *2* (5), 419–430.
(7) Drewes, G.; Bouwmeester, T. Global approaches to protein-protein interactions. *Curr. Opin. Cell Biol.* **2003**, *15* (2), 199–205.
(8) Ideker, T. A systems approach to discovering signaling and regulatory pathways - or, how to digest large interaction networks into relevant pieces. *Adv. Syst. Biol.* **2004**, *547*, 21–30.
(9) Noirot, P.; Noirot-Gros, M. F. Protein interaction networks in bacteria. *Curr. Opin. Microbiol.* **2004**, *7* (5), 505–512.
(10) Schachter, V. Protein-interaction networks: from experiments to analysis. *Drug Discovery Today* **2002**, *7* (11), S48–S54.
(11) Xia, Y.; Yu, H. Y.; Jansen, R.; Seringhaus, M.; Baxter, S.; Greenbaum, D.; Zhao, H. Y.; Gerstein, M. Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.* **2004**, *73*, 1051–1087.
(12) Chen, Y.; Xu, D. Computational analyses of high-throughput protein-protein interaction data. *Current Protein Pept. Sci.* **2003**, *4* (3), 159–180.
(13) Musso, G. A.; Zhang, Z.; Emili, A. Experimental and computational procedures for the assessment of protein complexes on a genome-wide scale. *Chem. Rev.* **2007**, *107* (8), 3585–3600.
(14) Fields, S.; Song, O. K. A novel genetic system to detect protein-protein interactions. *Nature* **1989**, *340*, 245–246.
(15) Rain, J. C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schachter, V.; Chemama, Y.; Labigne, A.; Legrain, P. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **2001**, *409* (6817), 211–215.
(16) Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M.; Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (8), 4569–4574.
(17) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M. J.; Johnston, M.; Fields, S.; Rothberg, J. M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **2000**, *403*, 623–627.
(18) Bertone, P.; Snyder, M. Advances in functional protein microarray technology. *FEBS J.* **2005**, *272* (21), 5400–5411.
(19) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. Global analysis of protein activities using proteome chips. *Science* **2001**, *293*, 2101–2105.
(20) Sidhu, S. S.; Fairbrother, W. J.; Deshayes, K. Exploring protein-protein interactions with phage display. *ChemBioChem* **2003**, *4* (1), 14–25.
(21) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002**, *415*, 141–147.
(22) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W. V.; Figeys, D.; Tyers, M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **2002**, *415*, 180–183.
(23) Arifuzzaman, M.; Maeda, M.; Itoh, A.; Nishikata, K.; Takita, C.; Saito, R.; Ara, T.; Nakahigashi, K.; Huang, H. C.; Hirai, A.; Tsuzuki, K.; Nakamura, S.; Altaf-Ul-Amin, M.; Oshima, T.; Baba, T.; Yamamoto, N.; Kawamura, T.; Ioka-Nakamichi, T.; Kitagawa, M.; Tomita, M.; Kanaya, S.; Wada, C.; Mori, H. Large-scale identification of protein-protein interaction of Escherichia coli K-12. *Genome Res.* **2006**, *16* (5), 686–691.
(24) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W. H.; Yang, X. C.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **2005**, *433* (7025), 531–537.
(25) Tackett, A. J.; DeGrasse, J. A.; Sekedat, M. D.; Oeffinger, M.; Rout, M. P.; Chait, B. T. I-DIRT, a general method for distinguishing

between specific and nonspecific protein interactions. *J. Proteome Res.* **2005**, *4* (5), 1752–1756.
(26) Frazier, M. E.; Johnson, G. M.; Thomassen, D. G.; Oliver, C. E.; Patrinos, A. Realizing the potential of the genome revolution: the genomes to life program. *Science* **2003**, *300* (5617), 290–293.
(27) Kovach, M. E.; Elzer, P. H.; Hill, D. S.; Robertson, G. T.; Farris, M. A.; Roop, R. M.; Peterson, K. M. 4 new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **1995**, *166* (1), 175–176.
(28) Landy, A. Dynamic, structural, and regulatory aspects of lambda-site-specific recombination. *Annu. Rev. Biochem.* **1989**, *58*, 913–949.
(29) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood, C. S. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. *Nat. Biotechnol.* **2004**, *22* (1), 55–61.
(30) Heidelberg, J. F.; Paulsen, I. T.; Nelson, K. E.; Gaidos, E. J.; Nelson, W. C.; Read, T. D.; Eisen, J. A.; Seshadri, R.; Ward, N.; Methe, B.; Clayton, R. A.; Meyer, T.; Tsapin, A.; Scott, J.; Beanan, M.; Brinkac, L.; Daugherty, S.; DeBoy, R. T.; Dodson, R. J.; Durkin, A. S.; Haft, D. H.; Kolonay, J. F.; Madupu, R.; Peterson, J. D.; Umayam, L. A.; White, O.; Wolf, A. M.; Vamathevan, J.; Weidman, J.; Impraim, M.; Lee, K.; Berry, K.; Lee, C.; Mueller, J.; Khouri, H.; Gill, J.; Utterback, T. R.; McDonald, L. A.; Feldblyum, T. V.; Smith, H. O.; Venter, J. C.; Nealson, K. H.; Fraser, C. M. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat. Biotechnol.* **2002**, *20* (11), 1118–1123.
(31) Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; ColladoVides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. The complete genome sequence of Escherichia coli K-12. *Science* **1997**, *277* (5331), 1453–1462.
(32) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **1999**, *17*, 1030–1032.
(33) Rey, F. E.; Oda, Y.; Harwood, C. S. Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. *J. Bacteriol.* **2006**, *188* (17), 6143–6152.
(34) Gorby, Y. A.; Yanina, S.; McLean, J. S.; Rosso, K. M.; Moyles, D.; Dohnalkova, A.; Beveridge, T. J.; Chang, I. S.; Kim, B. H.; Kim, K. S.; Culley, D. E.; Reed, S. B.; Romine, M. F.; Saffarini, D. A.; Hill, E. A.; Shi, L.; Elias, D. A.; Kennedy, D. W.; Pinchuk, G.; Watanabe, K.; Ishii, S.i.; Logan, B.; Nealson, K. H.; Fredrickson, J. K. Electrically conductive bacterial nanowires produced by Shewanella oneidensis strain MR-1 and other microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (30), 11358–11363.
(35) Kim, M. K.; Harwood, C. S. Regulation of Benzoate-CoA Ligase in Rhodopseudomonas palustris. *FEMS Microbiology Letters* **1991**, *83* (2), 199–203.
(36) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Spring Harbor, NY, 2001; pp1.84−1.117.
(37) Dehio, C.; Meyer, M. Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in Bartonella henselae following conjugal plasmid transfer from Escherichia coli. *J. Bacteriol.* **1997**, *179* (2), 538–540.
(38) Zeghouf, M.; Li, J.; Butland, G.; Borkowska, A.; Canadien, V.; Richards, D.; Beattie, B.; Emili, A.; Greenblatt, J. F. Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.* **2004**, *3* (3), 463–468.
(39) Hervey, W. J.; Strader, M. B.; Hurst, G. B. Comparison of digestion protocols for microgram quantities of enriched protein samples. *J. Proteome Res.* **2007**, *6* (8), 3054–3061.
(40) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
(41) Tabb, D. L.; McDonald, W. H.; Yates, J. R. DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1* (1), 21–26.
(42) Gattiker, A.; Michoud, K.; Rivoire, C.; Auchincloss, A. H.; Coudert, E.; Lima, T.; Kersey, P.; Pagni, M.; Sigrist, C. J. A.; Lachaize, C.; Veuthey, A. L.; Gasteiger, E.; Bairoch, A. Automated annotation of microbial proteomes in SWISS-PROT. *Comput. Biol. Chem.* **2003**, *27* (1), 49–58.
(43) Romine, M. F.; Elias, D. A.; Monroe, M. E.; Auberry, K.; Fang, R. H.; Fredrickson, J. K.; Anderson, G. A.; Smith, R. D.; Lipton, M. S. Validation of *Shewanella oneidensis* MR-1 small proteins by AMT tag-based proteome analysis. *OMICS* **2004**, *8* (3), 239–254.
(44) Sharp, J. L.; Anderson, K. K.; Hurst, G. B.; Daly, D. S.; Pelletier, D. A.; Cannon, W. R.; Auberry, D. L.; Schmoyer, D. D.; McDonald, W. H.; White, A. M.; Hooker, B. S.; Victry, K. D.; Buchanan, M. V.; Kery, V.; Wiley, H. S. Statistically Inferring Protein-Protein Associations with Affinity Isolation LC-MS/MS Assays. *J. Proteome Res.* **2007**, *6* (9), 3788–3795.
(45) Edwards, A. M.; Kus, B.; Jansen, R.; Greenbaum, D.; Greenblatt, J.; Gerstein, M. Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* **2002**, *18* (10), 529–536.
(46) Hiratsu, K.; Amemura, M.; Nashimoto, H.; Shinagawa, H.; Makino, K. The *rpoE* gene of *Escherichia coli*, which encodes sigma(E), is essential for bacterial-growth at high-temperature. *J. Bacteriol.* **1995**, *177* (10), 2918–2922.
(47) Liu, H. B.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201.
(48) von Mering, C.; Jensen, L. J.; Kuhn, M.; Chaffron, S.; Doerks, T.; Kruger, B.; Snel, B.; Bork, P. STRING 7−recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **2007**, *35* (Suppl. 1), D358–D362.
(49) Bowers, P. M.; Pellegrini, M.; Thompson, M. J.; Fierro, J.; Yeates, T. O.; Eisenberg, D. Prolinks: a database of protein functional linkages derived from coevolution. *GenomeBiology* **2004**, *5* (5), R35.
(50) Mah, T. F.; Kuznedelov, K.; Mushegian, A.; Severinov, K.; Greenblatt, J. The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes Dev.* **2000**, *14* (20), 2664–2675.
(51) Rouby, J.; Pugniere, M.; Mani, J. C.; Granier, C.; Monmouton, P.; Saint Germain, S. T.; Leonetti, J. P. Characterization of monoclonal antibodies against *Escherichia coli* core RNA polymerase. *Biochem. J.* **2002**, *361*, 347–354.
(52) Banerjee, S.; Chalissery, J.; Bandey, I.; Sen, R. J. Rho-dependent transcription termination: More questions than answers. *J. Microbiol.* **2006**, *44* (1), 11–22.
(53) Gamer, J.; Bujard, H.; Bukau, B. Physical interaction between heat-shock proteins DnaK, DnaJ, and GrpE and the bacterial heat-shock transcription factor-sigma(32). *Cell* **1992**, *69* (5), 833–842.
(54) Venkateswaran, K.; Moser, D. P.; Dollhopf, M. E.; Lies, D. P.; Saffarini, D. A.; MacGregor, B. J.; Ringelberg, D. B.; White, D. C.; Nishijima, M.; Sano, H.; Burghardt, J.; Stackebrandt, E.; Nealson, K. H. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int. J. Syst. Bacteriol.* **1999**, *49* (2), 705–724.

PR8001832