A high-throughput immobilized bead screen for stable proteins and multi-protein complexes

Meghan A.Lockard[1], Pawel Listwan[1], Jean-Denis Pedelacq[1,2,3], Stéphanie Cabantous[1,4,5,6], Hau B.Nguyen[1], Thomas C.Terwilliger[1] and Geoffrey S.Waldo[1,7]

[1]Biosciences Division, MS-M888, Los Alamos National Laboratory, PO Box 1663, Los Alamos, NM 87545, USA, [2]CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), 205 route de Narbonne, F-31077 Toulouse, France, [3]Université de Toulouse, UPS, IPBS, F-31077 Toulouse, France, [4]INSERM UMR1037-Cancer Research Center of Toulouse, [5]Université de Toulouse and [6]Institut Claudius Regaud, F-31052 Toulouse, France

[7]To whom correspondence should be addressed.
E-mail: waldo@lanl.gov

Edited by Todd Yeates

We describe an *in vitro* colony screen to identify *Escherichia coli* expressing soluble proteins and stable, assembled multiprotein complexes. Proteins with an N-terminal 6His tag and C-terminal green fluorescent protein (GFP) S11 tag are fluorescently labeled in cells by complementation with a coexpressed GFP 1–10 fragment. After partial colony lysis, the fluorescent soluble proteins or complexes diffuse through a supporting filtration membrane and are captured on Talon® resin metal affinity beads immobilized in agarose. Images of the fluorescent colonies convey total expression and the level of fluorescence bound to the beads indicates how much protein is soluble. Both pieces of information can be used together when selecting clones. After the assay, colonies can be picked and propagated, eliminating the need to make replica plates. We used the method to screen a DNA fragment library of the human protein p85 and preferentially obtained clones expressing the full-length 'breakpoint cluster region-homology' and NSH2 domains. The assay also distinguished clones expressing stable multi-protein complexes from those that are unstable due to missing subunits. Clones expressing stable, intact heterotrimeric *E.coli* YheNML complexes were readily identified in libraries dominated by complexes of YheML missing the N subunit.
Keywords: chemical lysis/high-throughput screening/IMAC beads/protein tagging/split GFP

Introduction

A common problem in molecular biology is the misfolding and aggregation of proteins expressed in heterologous hosts.

A powerful solution is to screen libraries of mutants or domains for more soluble and stable forms. Many approaches of this kind have been developed, but each has significant drawbacks. Some approaches accurately measure *in vitro* soluble protein, but lack throughput (Knaust and Nordlund, 2001). Recently developed methods have higher throughput, but require unique and often expensive equipment (Reich *et al*., 2006; Tarendeau *et al*., 2007, 2008; Angelini *et al*., 2009). Colony-based screens (Cornvik *et al*., 2005; Cabantous and Waldo, 2006) are some of the more promising methods to balance expense and throughput. The split green fluorescent protein (GFP) (Cabantous *et al*., 2005) solubility screen operates at the single colony level in living bacterial colonies, facilitates the recovery of viable clones and specifically labels proteins with fluorescence using a small, easily detectable, non-perturbing tag. The split GFP screen has been used to screen large libraries for soluble proteins guided by the fluorescent signal produced by each colony (Cabantous and Waldo, 2006). However, unstable proteins that fold using chaperones can become less soluble after lysis and may be incorrectly designated by such screens (Fujiwara *et al*., 2010). As a result, the split GFP colony assay requires follow-up liquid cultures and confirmatory *in vitro* complementation assays of promising clones in multi-well plates (Cabantous and Waldo, 2006; Listwan *et al*., 2009). It would be desirable to measure the solubility of proteins *in vitro* after lysis, with the same throughput as the colony-based assay.

One important field in which soluble expression of proteins is particularly important is structural biology. To be useful in structure determination, target proteins or protein complexes must be soluble and sufficiently stable for purification and crystallization or NMR studies (Terwilliger, 2004; Graslund *et al*., 2008a; Terwilliger *et al*., 2009). Proteins with high solubility often can be purified more readily than those that are sparingly soluble (Graslund *et al*., 2008a). Consequently, an assay that reports directly on solubility would be of substantial use. The fraction of a protein expressed in a soluble form is typically assessed from the measured soluble expression divided by the total expression (Graslund *et al*., 2008a,b). When carried out in intact cells, the split GFP assay described above reports either total protein or soluble protein as cellular fluorescence depending on the order in which the GFP S11-tagged protein and GFP 1–10 detector fragment are expressed (Cabantous and Waldo, 2006). The *in vivo* split GFP assay cannot simultaneously report both parameters in the same experiment. A colony-based assay that reports soluble and total protein in the same experiment could aid structural biologists in the selection of promising crystallization or NMR targets from hundreds or even thousands of trial constructs.

A specific application of high-throughput, high-content solubility screens is to find soluble domains of large, multi-domain proteins, especially for structural determination. For example, Hart and coworkers have developed expression of soluble protein by random incremental truncation in which random incremental truncation libraries are grown in multi-well plates, micro-colonies arrayed onto membranes using spotting pins and screened for soluble expression by detecting a tag (Yumerefendi et al., 2010). This protocol requires some expensive specialized equipment, but has led to spectacular successes (Tarendeau et al., 2007, 2008). GFP 'folding' reporters (Waldo et al., 1999) have also been used to this end previously (Kawasaki and Inagaki, 2001) leading to the successful structural determination of domains from larger proteins, such as the N-terminal domain of the telomerase reverse transcriptase (Jacobs et al., 2005, 2006). Although simple to use in colony assays, a potential limitation of using folding reporters for soluble domain screens is that they do not directly measure the solubility of the test protein. Rather, the GFP folding reporter places a GFP downstream of the protein of interest. Misfolding of the upstream protein is thought to interfere with the subsequent folding of the GFP. Proteins leading to bright GFP fusions tend to be better folded and more soluble when expressed alone, compared with proteins that produce faint GFP fusions. When using folding reporters, domains slowly aggregating after the reporter is committed to fold could lead to false positives. This problem was the impetus for developing the split GFP system (Cabantous et al., 2005) that directly measures soluble protein using a fused C-terminal GFP strand 11 tag. Regardless of the format, reporter systems using only a single C-terminal tag can give false positives due to upstream internal ribosome initiation sites (Cabantous et al., 2008). By using an additional N-terminal reporter, such as a 6His tag, in addition to a C-terminal reporter, one can ensure that the selected fragments have both ends of the target intact. The in vitro colony screen described here utilizes both the split GFP system to accurately report solubility, as well as an N-terminal 6His tag, to exclusively screen for fragments with both termini intact, making it an ideal tool for the application of mapping soluble domain fragments.

Protein complexes are especially difficult for structure determination. Complexes must be soluble and remain stably assembled throughout the production and crystallization pipeline (Terwilliger, 2004; Graslund et al., 2008a; Terwilliger et al., 2009). When the members of a complex are obligatory folding partners and are insoluble when expressed alone, an easy way to screen for stable expression of the full-length complex is to monitor solubility of one or more of the member subunits. For example, the folding and assembly of the *Escherichia coli* integration host factor heterodimer complex (Wang and Chong, 2003) was monitored with a GFP tag (Waldo et al., 1999; Waldo, 2003a,b). The GFP was attached to one of the subunits whose folding and solubility depended on the coexpression of the other subunit. Since the assay readout is cellular fluorescence, large libraries of peptides could be screened to find obligatory folding partners (Wang and Chong, 2003). This strategy cannot be used, however, if the subunits are soluble when expressed separately. Measuring the solubility of a single-protein subunit of a complex using any single-tag system,

including the split GFP, does not indicate that the complex is stable enough for copurification (Strong et al., 2006; Graslund et al., 2008a). These issues could be addressed by attaching an affinity tag to one member of a complex, and a second detection tag to a different subunit. The detection of the tag after the complex has been directionally captured on affinity media could indicate that the complex is assembled and stable. Such approach has proved useful, but current implementations for structural biology remain low throughput (Graslund et al., 2008a). A colony-based assay giving similar information would be useful for screening large collections of clones for stable complexes.

Here we describe an assay that overcomes most limitations of current methods. Colonies expressing target proteins tagged with GFP strand 11 are complemented with the reporter GFP 1–10. Some of the cells in the colony are lysed and subjected to a rigorous solubility and stability screen that combines filtration, diffusion through agarose and capture or binding to immobilized metal affinity resin. In the course of the assay, both total protein expression and soluble protein are measured and used to guide the identification of desired clones, allowing an estimate of the fraction of the total expressed protein that is soluble (Fig. 1). The lysis method leaves enough viable cells in the colony for subsequent picking without the need for replica plates. Finally, because the method uses the 'tandem affinity purification' concept (Puig et al., 2001)—two orthogonal tags where one tag is for in vitro capture and another is for detection—the split GFP technology can be applied to a broad scope of biological problems, including the identification of stably assembled multiprotein complexes (Rigaut et al., 1999).
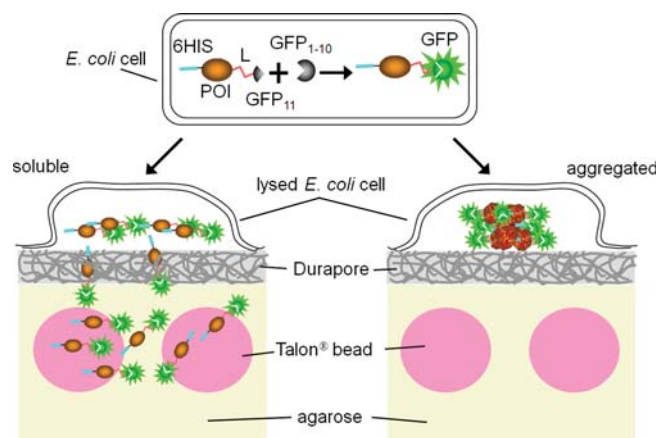

Fig. 1. Principle of bead-binding assay for soluble proteins. A protein of interest is flanked by a 6His tag on its N-terminus and the GFP S11 tag (β-strand 11, residues 215–230) on its C-terminus via a flexible linker (L). The complementary GFP 1–10 detector fragment (β-strands 1–10, residues 1–214) is coexpressed separately in the same cell (*top*). The GFP S11 tag rapidly associates with the GFP 1–10, committing the GFP to fold and form the fluorophore, whether the protein of interest subsequently remains soluble (*left*) or aggregates (*right*). Colonies resting on Durapore® membranes are partially lysed, soluble fluorescently labeled proteins diffuse through the Durapore® membrane and bind to Talon® resin affinity beads immobilized in agarose (*left*). Insoluble aggregates cannot pass through the Durapore®, or the agarose matrix (*right*).

Table I. Eighteen control proteins from *P.aerophilum*

#[e]	Protein[f]	Protein expressed alone[a]			Lysis in TNG buffer[b]				Lysis in Solulyze®[c]				Immobilized bead assay[d]	
		$F_{sol}^{g,h}$	$F_{pel}^{g,i}$	Fraction soluble[j]	$F_{sol}^{g,k,l}$	$F_{pel}^{g,m,n}$	Fraction soluble[j]	$F_{beads}^{g,o,p}$	$F_{sol}^{g,k,q}$	$F_{pel}^{g,m,r}$	Fraction soluble[j]	$F_{beads}^{g,o,s}$	Colony[g,t]	$F_{beads}^{g,u}$
1	DNA-directed RNA polymerase	18 955	4860	0.80	41 365	16 130	0.72	17 630	47 725	4110	0.92	14 000	147	144
2	Sulfite reductase (dissimilatory subunit)	26 510	0	1.00	59 730	2860	0.95	21 005	62 890	1320	0.98	13 485	126	134
3	c-Type cytochrome biogenesis factor	40 855	19 365	0.68	62 445	9515	0.87	25 015	70 990	11 815	0.86	13 610	148	118
4	Translation initiation factor	30 615	110	1.00	45 235	6505	0.87	16 410	61 900	6450	0.91	17 170	152	154
5	Ribosomal protein S9p	70 700	490	0.99	53 330	1145	0.98	19 300	64 215	2000	0.97	11 750	130	119
6	Polysulfide reductase subunit	865	10 340	0.08	0	5455	0.00	0	1115	6145	0.15	95	35	2
7	Nucleoside diphosphate kinase	3415	23 345	0.13	17 035	30 640	0.36	5685	18 270	25 275	0.42	4025	123	27
8	Tartrate dehydratase β-subunit	775	24 975	0.03	4855	19 130	0.20	1625	3405	21 640	0.14	345	95	2
9	3-Hexulose 6-phosphate synthase	31 650	515	0.98	28 785	6730	0.81	11 620	38 000	7035	0.84	13 895	116	82
10	Hydrogenase formation hypE	18 885	21 790	0.46	19 135	31 070	0.38	7845	29 590	24 485	0.55	8815	149	94
11	Methyltransferase	2170	5080	0.30	13 995	11 175	0.56	6430	4435	24 925	0.15	420	111	15
12	Chorismate mutase	15 405	1750	0.90	25 995	1775	0.94	9860	29 165	3090	0.90	6255	104	84
13	Tyrosine t-RNA synthetase	33 395	185	0.99	28 780	2800	0.91	11 200	37 735	3810	0.91	9285	119	117
14	nirD protein	8955	90	0.99	13 440	3325	0.80	3210	18 440	2900	0.86	1655	71	51
15	Soluble hydrogenase	0	3325	0.00	755	12 745	0.06	105	1000	15 575	0.06	25	74	0
16	Aspartate aldehyde dehydrogenase	380	3455	0.10	315	14 290	0.02	85	0	12 975	0.00	40	77	0
17	Phosphate cyclase	28 885	1760	0.94	22 280	7435	0.75	9420	27 560	8930	0.76	8675	98	88
18	Purine-nucleoside phosphorylase	1035	2575	0.29	2170	14 855	0.13	75	3590	16 655	0.18	0	93	4

[a]Protein with S11 tag expressed alone in *E.coli* liquid shake cultures from the pTET promoter plasmid (Cabantous *et al.*, 2005).
[b]Tagged protein coexpressed with GFP 1–10 and sonicated in TNG buffer.
[c]Coexpressed with GFP 1–10 then sonicated in SoluLyse® buffer (Materials and methods).
[d]Coexpressed with GFP 1–10 as colonies on membranes (Materials and methods).
[e]Number of indicated test protein.
[f]Indicated test protein cloned from *P.aerophilum* by PCR as previously described (Waldo *et al.*, 1999; Cabantous *et al.*, 2005).
[g]Relative uncertainty of indicated measurement is ∼5%.
[h]Fluorescence of soluble protein measured by plate reader after complementation of soluble fraction by GFP 1–10 *in vitro*. Background of 1485 subtracted.
[i]Fluorescence after *in vitro* complementation of urea-solubilized pellet fraction using GFP 1–10. Background of 60 subtracted.
[j]Fraction soluble = $F_{sol}/(F_{sol} + F_{pel})$ of corresponding fractions.
[k]Fluorescence of soluble protein measured by plate reader.
[l]Background of 3790 subtracted.
[m]Fluorescence of resuspended pellet fraction.
[n]Background of 465 subtracted.
[o]Fluorescence of Talon® resin beads after binding of indicated soluble fraction in batch mode.
[p]Background of 785 subtracted.
[q]Background of 3000 subtracted.
[r]Background of 460 subtracted.
[s]Background of 385 subtracted.
[t]Mean fluorescence of three intact *E.coli* colonies for indicated construct, measured using a digital camera. Background of 30 subtracted.
[u]Mean fluorescence of protein bound to Talon® resin beads, released from corresponding partially lysed colony (Materials and methods). Background of 20 subtracted.

Materials and methods

Cloning and expression of control proteins, complexes and DsRed control

Eighteen control proteins from *Pyrobaculum aerophilum* (Table I) were cloned into the N6His pTET S11 SpecR ColE1 ORI vector and transformed into BL21 (DE3) bearing the pET GFP 1–10 KanR p15 ORI vector as described (Cabantous *et al.*, 2005). Polycistrons encoding test complexes (Table II) were cloned by polymerase chain reaction (PCR) from genomic DNA in the N6His pTET S11 vector (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006) using the specified primers (Supplementary Table S1) as previously described (Cabantous *et al.*, 2005). Translation of the first protein in the polycistron begins at the vector ribosome-binding site, while other protein subunits are translated from their native ribosome-

Table II. Multi protein complexes and trial constructs

#	Trial construct	Description	Source organism	Oligomerization
1	Yhe NML[a]	Sulfur transfer relay	*E.coli*	Trimer[a]
2	Yhe ML[b]	Sulfur transfer relay	*E.coli*	Unstable (expected)[b]
3	Rv2431c/ Rv2430c[c]	PE/PPE	*M.tuberculosis*	Dimer[c]
4	Rv2430[d]	PE/PPE	*M.tuberculosis*	Unstable (expected)[d]
5	Rv0264/ Rv0263[e]	Allophanate hydrolase	*M.tuberculosis*	Dimer
6	YBGJ/ YBJK[e]	Allophanate hydrolase	*E.coli*	Dimer
7	RFA 2683/ 2682[e]	Allophanate hydrolase	*Rhodopseudomonas palustris*	Dimer
8	0428/ 0427[e,f]	Allophanate hydrolase	*M.smegmatis*	Dimer[e,f]

[a]Tus DCB (Yhe NML) polycistronic trimer recently solved (Numata *et al.*, 2006). Positive control.
[b]Considerable buried surface involved in the contacts between YheN and the adjacent YheM and YheL subunits (Numata *et al.*, 2006), omission of YheN is expected to destabilize the complex. Negative control.
[c]Bicistronic PE/PPE dimeric complex from *M.tuberculosis* whose structure was recently solved (Strong *et al.*, 2006).
[d]Based on the buried surface between Rv2431 and Rv2430 (Strong *et al.*, 2006), omitting Rv2431 should likely lead to stability problems or aggregation.
[e]Predicted bicistronic allophanate hydrolase.
[f]Protein bicistron recently expressed and purified as dimer; structure solved as stable dimer (David Eisenberg and coworkers, M. Kaufmann, personal communication, DOI:10.2210/pdb3mml/pdb).

binding sites in the polycistron. For red fluorescent marker clones, GFP 1–10 was replaced with DsRED (Clontech) and cotransformed with the empty N6His pTET S11 vector. Clones were plated, grown and coinduced as previously described (Cabantous *et al.*, 2005) but a hydrophilic 0.45 μm Durapore® (Millipore) membrane was used in place of nitrocellulose to reduce adventitious protein binding during lysis steps in subsequent experiments.

Optimization of cell lysis conditions for measuring protein solubility

An important goal of our approach is to have an assay that yields estimates of solubility that are as close as possible to those obtained by sonication of cells expressing a protein followed by measurement of the amount of the protein that is soluble and insoluble. We carried out a series of tests to identify methods for cell lysis that maximized the correlation between the results of our assays and results from sonication and measurement of fractional solubility of a series of 18 test proteins (Waldo *et al.*, 1999; Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). In initial tests, SoluLyse® was used to lyse colonies expressing the 18 control proteins, and TNG was the buffer used in the capture plate (Supplementary Fig. S1, *column 2*). Control proteins #7 and #7 were less soluble compared with sonication in TNG (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). Next, we tested four additional conditions: two commercial lysis buffers,

SoluLyse® (Genlantis) or Bugbuster® (Novagen EMD); and two capture plate buffers (TNG or Tris HCl) (Supplementary Fig. S1). The solubility of control protein #5 decreased compared with sonication when lysed with Bugbuster® and captured using Talon® plates with Tris HCl. Control proteins #7 and #9 were most soluble and behaved most similarly to sonication when lysed with SoluLyse® in Tris HCl and captured using Talon® plates with Tris HCl (Supplementary Fig. S1, columns 3, 4). A partial factorial screen (Armstrong *et al.*, 1999) was used to test the effects of five chemical adjuvants on the solubility of control protein #9 during chemical lysis (Supplementary Table S2). We selected protein #9 as the test candidate since its behavior differed most from sonication in tests on the 18 control proteins in the colony-based assay (Supplementary Fig. S1, column 2). The five factors and the two states of each factor to be tested included: included SoluLyse® (75 or 35% v/v); NaCl (0.15 M or none); Tris-HCl pH 7.5 (100 or 20 mM); glycerol (10% v/v or none); and 10 AU/ml Benzonase® (Sigma) with 2 mM of the cofactor $MgCl_2$ (or no Benzonase®/$MgCl_2$). Five factors and two states of each yielded a partial factorial screen with 32 conditions. A 200 μl aliquot of each lysis cocktail (Supplementary Table S2) was placed in a well of a 96-well PCR plate. Three 3.5 ml cultures of control protein #9 were grown, coinduced and pooled. Two hundred microliters of aliquots were placed in 32 wells of a second 96-well PCR plate, centrifuged and then the pellets were lysed by adding the array of the lysis cocktails. Once the lysis cocktails were added, cells were mixed using the Biomek FX liquid handling robot (Beckman–Coulter) for 10 min at room temperature. The samples were then moved to a Rotanta 46 RSC refrigerated robotic centrifuge (Hettich AG) and centrifuged for 20 min at 4000 r.p.m. Following the centrifugation, the plates were again returned to Biomek FX deck, 120 μl of supernatant (soluble fraction) was aspirated from the wells without disturbing the pellet and transferred to a fresh plate. Pellets were dried, washed with 200 μl of TNG and then resuspended in 200 μl of TNG, and transferred to a fresh plate, also using the Biomek FX. The fluorescence values of both plates were subsequently read on a DTX plate reader (Beckman–Coulter) (Listwan *et al.*, 2009). Factors that resulted in solubility most similar to sonication were identified by principal component analysis (Armstrong *et al.*, 1999). Benzonase® and $MgCl_2$ strongly increased Talon®-bound soluble protein, whereas NaCl and glycerol decreased it (Supplementary Fig. S1 and Table S2). The final optimal chemical lysis buffer, which resulted in solubility most similar to sonication, included SoluLyse® 75% v/v, 100 mM Tris-HCl pH 7.5, 10 AU/ml Benzonase® and 2 mM $MgCl_2$.

Optimization of lysis conditions to measure the stability of protein complexes

A second important goal of our approach is to measure the formation of protein complexes. In order to be useful, our assay would need to capture complexes that were previously reported as stable *in vitro*, including YheNML (Numata *et al.*, 2006), Rv2431c/2430c (Strong *et al.*, 2006) and the heterodimer allophanate hydrolase from *Mycobacterium smegmatis* Rv0264/Rv0263 whose structure was recently solved by David Eisenberg and coworkers (David Eisenberg and coworkers, M. Kaufmann, personal communication, DOI:10.2210/pdb3mml/pdb) (Table II, Supplementary

Fig. S2a). The complex assembly screen was first tested on capture plates made with Tris HCl to see how stable the complexes were without additional additives. Only when Talon® capture plates included 100 mM Tris HCl, 150 mM NaCl, 10% glycerol v/v (TNG) did the YheNML, Rv2431c/2430c and Rv0264/0263 yield bound fluorescent spots as expected (Supplementary Fig. S2b). The inclusion of NaCl and glycerol in the capture plates appeared to stabilize these complexes relative to Tris HCl alone.

Growth of liquid cultures and assay of soluble and insoluble protein in vitro

Control proteins were expressed alone and assayed for soluble and insoluble protein *in vitro* by complementation with the GFP 1–10 reagent (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). For coexpression experiments in liquid culture, overnight Luria Bertani (LB) cultures expressing the 18 control proteins were diluted 100-fold into 3.5 ml LB, shaken at 350 r.p.m. for 2 h at 37°C, induced with 350 ng/ml anhydrotetracycline (ANTET) and 1 mM isopropylthiogalactoside (IPTG) for 4 h at 37°C, 350 r.p.m. Cultures were diluted to 0.4 $OD_{600\,nm}$ with LB. Three milliliters of each culture was centrifuged for 5 min (16 000 × g), the cell pellet was suspended in 200 μl of TNG, and sonicated 3 times for 20 s (Branson Sonifier, 50% duty cycle, centrifuged 2 min between each sonication cycle). The sonicated samples were centrifuged for 10 min (16 000 × g), and the ∼200 μl supernatant (soluble fraction) was recovered by pipetting. The pellet fraction was washed twice with 200 μl of TNG, and resuspended by brief sonication in 200 μl TNG. Forty microliters of the soluble or pellet fraction was suspended with 160 μl of TNG and the fluorescence was measured (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). In a separate experiment, 40 μl of the remaining soluble fraction was incubated with 40 μl of Talon® (Clontech) (40 μl bed volume) for 10 min at 22°C, washed three times with 200 μl of TNG, suspended with 160 μl TNG and the fluorescence was measured (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). For tests with Sololulyze®, samples were treated as above except for 600 μl of the optimized SoluLyse® cocktail (Materials and methods, *Optimization of cell lysis conditions for measuring protein solubility*) was used in place of TNG for sonication. Hundred and twenty microliters of the samples were used to compensate for the 3-fold dilution caused by the larger lysis buffer volume. Backgrounds subtracted varied according to the experiment (Table I). Soluble fraction values had predicted higher backgrounds (e.g. 3790) due to the presence of soluble, autofluorescent flavins, the higher concentration of ANTET, etc. in the lysates. Pellet fractions, which do not have these autofluorescent species and go through a wash step to remove residual ANTET, had lower backgrounds (e.g. 60).

Talon® bead-blot assay plates

Silicone vacuum grease (Dow Corning) was applied using a gloved finger to the inner walls of two 150 mm Bauer plates (Fisher, see Supplementary Fig. S3). The interior base of each plate was protected from grease using a custom-made metal shield as depicted in the Supplementary Fig. S3. Thirty microliters of 50% v/v Talon® resin ethanol slurry was placed in a 50 ml Falcon® (Beckman Dickinson) tube and centrifuged at 500 × g for 5 min (Beckman J2-21), the supernatant was discarded, and the beads were washed three

times with 20 ml of TNG buffer. The supernatant was discarded, and the 50 ml Falcon® containing the washed resin bed was preheated in a water bath (80°C, 5 min). Three grams of agarose (Invitrogen) was suspended and dissolved in either 200 ml of 100 mM Tris HCl, pH 7.5 (for control proteins and control protein libraries), or in 200 ml of TNG (for multi-protein complexes). The molten agarose was poured into the Talon® resin to a final volume of 50 ml (∼15 ml bed volume of beads and ∼35 ml agarose), gently mixed by inverting the tube, and 25 ml of the suspension poured into each of the prepared Bauer plates. The slabs were misted with ethanol to remove bubbles prior to gelling, solidified by cooling ∼5 min, overlaid with ∼50 ml of molten agarose in the appropriate buffer (TNG or Tris HCl) and then the slabs were cooled to solidify (20 min). The agarose slabs were placed onto plastic wrap bead-side up, 5 ml of buffer was added to the empty Bauer plates, and the agarose slabs were replaced bead-side up, seated by striking on a paper towel stack, the displaced supernatant discarded, then the plates were dried for ∼1 h in a laminar flow hood prior to wrapping in plastic film for storage face up at 4°C for up to 3 days.

Colony induction, imaging and partial lysis of clones on capture plate

Freezer stocks of 1.0 OD were diluted, plated to obtain well-spaced single colonies (Fig. 2a) and coinduced as described (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006) except that Durapore® membranes were used *in lieu* of nitrocellulose. The membrane with the induced colonies was moved onto the capture plate, illuminated using an Illumatool Lighting System (LightTools Research) equipped with a 488 nm excitation filter and photographed with a digital camera (Olympus Camedia C-5060) through a 510 nm long pass glass filter (LightTools Research) to record fluorescence proportional to total protein expression. A fine-mist spray bottle was filled with ∼25 ml of the appropriate lysis cocktail, sufficient reagent was misted to just wet the surface of the colonies and membrane (∼2 ml). The plate was dried for ∼2 min, and the misting was repeated three additional times. Multiple repetitions of the misting ensure that the colonies on the plate are lysed uniformly. The plate was allowed to incubate at 37°C for 2 h, and then the membrane bearing the partially lysed colonies was returned to the original LB agar plate. The capture plate was imaged (Cabantous *et al.*, 2005) with exposures up to 4 s to record faint Talon® spots. Investigators can expect screening, image analysis and selection to take up to 2 days.

Image analysis to identify desired clones, image projection and 'guided' colony picking

Photographs of colony membranes and Talon® bead capture plates were taken with a digital camera (Olympus Camedia C-5060) under the same spectral conditions as the colonies before lysis (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). Images of the colonies and capture plate were manually aligned based on the DsRED spots using the public domain software 'HDR Alignment Plug-in' for ImageJ® (Wayne Rasband, NIH). Using the aligned capture plate image, compact fluorescent spots were identified and the corresponding colonies were marked with a white dot in the colony image using Paint Shop Pro® (JASC Software). Marked
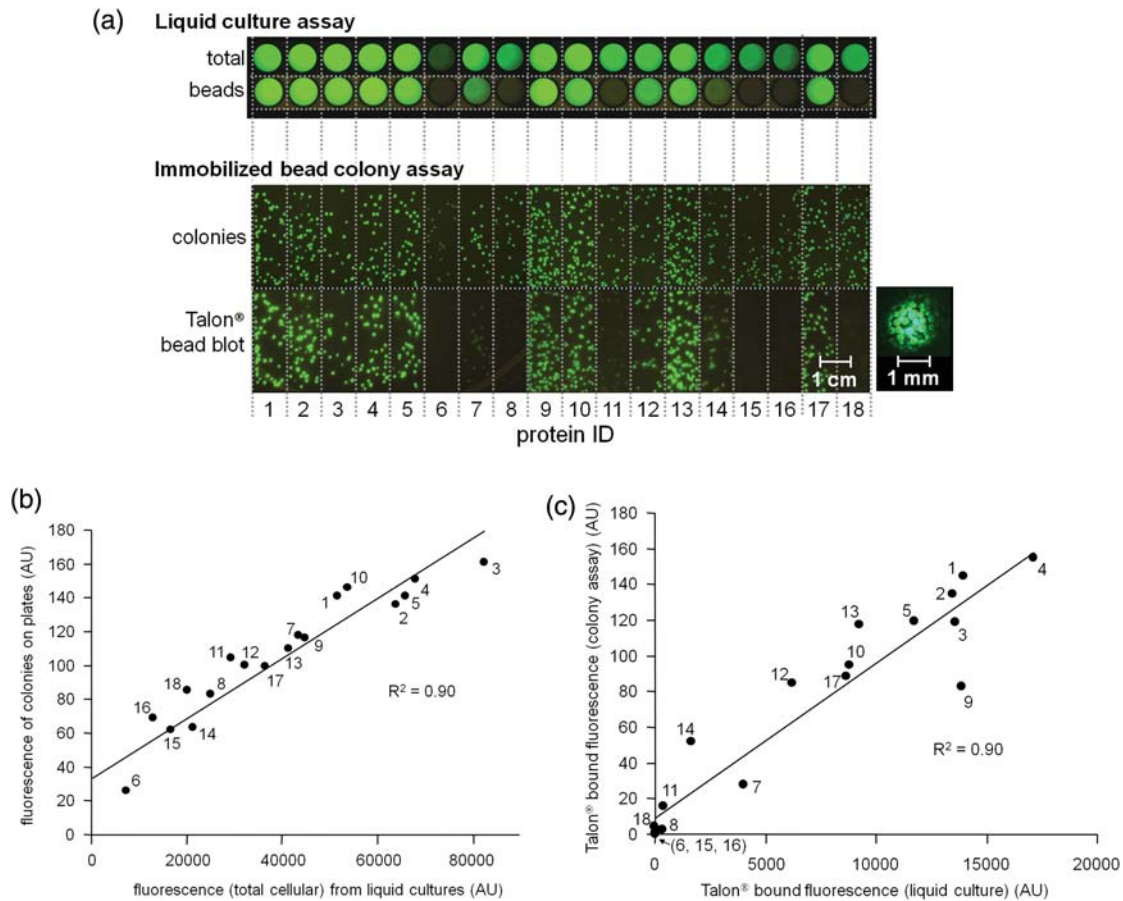
Fig. 2. Comparison of liquid culture and immobilized bead assays for total expressed protein and bead-bound soluble protein using GFP fluorescence. (**a**) Application of the GFP colony filtration immobilized bead assay outlined in Fig. 1 to 18 control proteins from *P.aerophilum* (Table I). Microtiter plate wells for control proteins expressed in 3 ml LB liquid culture. Total protein imaged at $\frac{1}{2}$ s exposure (*first row*) and the soluble bead-bound protein imaged at 2 s exposure (*second row*). Image of *E.coli* colonies expressing tagged proteins coexpressed with the GFP 1–10 detector fragment at 37°C for 4 h, reflecting total expression ($\frac{1}{2}$ s exposure) (*third row*). Image of the capture plate under the *E.coli* colonies expressing tagged proteins after lysis via repetitive misting of the SoluLyse® cocktail (4 s exposure) (*fourth row*). Illumatool® (Light Tools Research) used with 488 nm excitation, photographed through 520 nm-long pass filter. Ten-fold magnification of Talon® resin blot for a single colony of control protein #2 (*inset*) was imaged by a confocal microscope (Olympus IX-81) and a CCD camera (Olympus DP71) showing individual Talon® beads with bound green fluorescence. (**b**) Plot of coinduction fluorescence (proportional to total expression) of 18 control proteins in Table I. Fluorescence measured for single colonies (*y-axis*) or as liquid culture fluorescence (*x-axis*). See Table I for tabulated values. Plot is a graphical representation of the corresponding fractions in the photographs from (a) above. (**c**) Comparison of Talon®-bound protein in the colony-based immobilized bead assay (*y-axis*) and the Talon® bound in batch mode from liquid culture lysates (*x-axis*). Data are tabulated in Table I. Plot is a graphical representation of the corresponding fractions in (a).

colonies were given reference numbers by using the object finding function in UTHSCSA Image Tool® 3.0 (Wilcox, Drove, McDavid and Greer). The highlighted and numbered colony image was then displayed using Paint Shop Pro, and projected onto the original cell colonies using an MPro110® microprojector (3 M). Superimposition of the image and target was optimized by adjusting the projection distance (~30 cm) and by adjusting the magnification in Paint Shop Pro® (~34% full-scale setting) using the DsRED colonies as a guide. The desired colonies were easily identified by white spots from the projector. Colonies were picked using sterile toothpicks, and grown in 3 ml of LB and antibiotics overnight at 32°C. Plasmids were prepared using a Qiaprep® plasmid purification kit (Qiagen), and 20% v/v glycerol freezer stocks were stored at −80°C. To estimate total expressed protein in colonies and the protein captured on Talon® beads, images of colonies and corresponding Talon® blots were analyzed for average integrated fluorescence using Imagetool® (Scion). For each of the 18 control proteins, three colonies of similar size were analyzed and their values were averaged and then

compared reader measurements of Talon® batch-binding from liquid culture trials (Fig. 2b and c).

p85 large fragment library, ORF selection and cloning

Preparation of the open reading frame (ORF) fragment library of *Homo sapiens* phosphoinositide-3-kinase, p85 and selection of fragments without stop codons using a fused *E.coli* dihydrofolate reductase (DHFR) will be described in greater detail in a forthcoming publication (Pédelacq *et al.*, under peer review). Briefly, the p85 ORF was amplified by PCR from plasmid PIK3R1, NM_181523 (Origene), digested with DNAse-I and resolved by agarose gel electrophoresis. The DNA was visualized by ethidium bromide staining, then a slab of gel containing DNA fragments 400–800 bp was excised and recovered with a QIAquick® gel extraction kit (Qiagen). The DNA fragment library was blunt cloned into an internal permissive site of *E.coli* DHFR, and plated on agar plates with 6 μg/ml trimethoprim (TMP) as described (USPTO 7390640) to remove fragments with stop codons. Overnight colonies from lawns of ~10^6 clones, estimated by
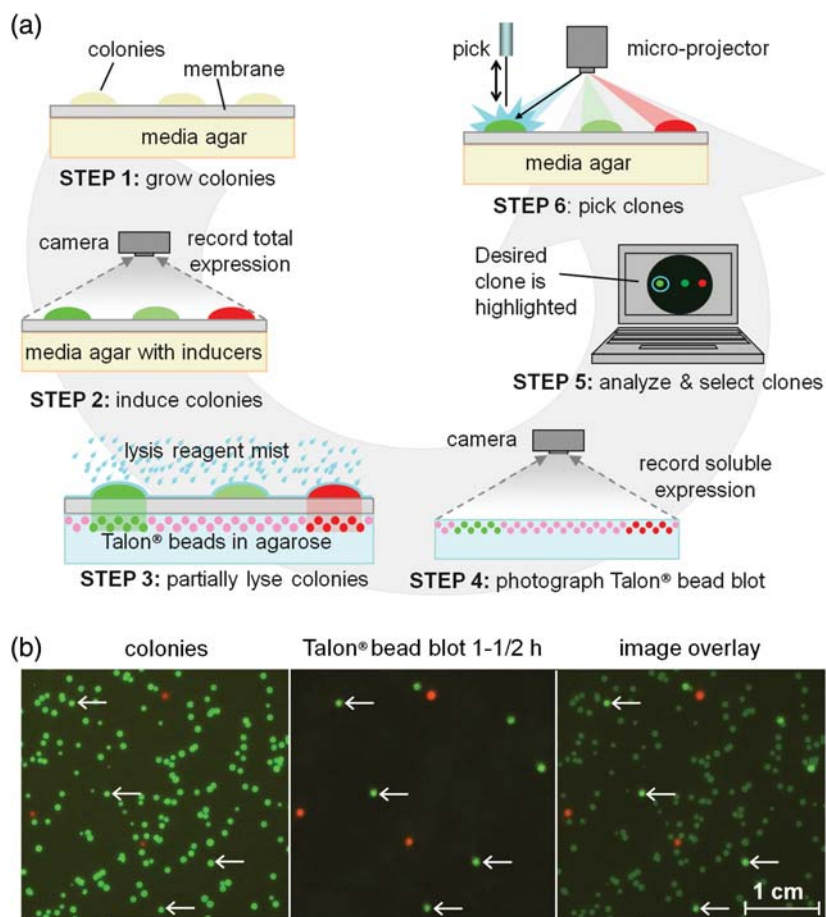
Fig. 3. (**a**) Overview of the split GFP colony filtration assay applied to colony picking. *Step 1: Escherichia coli* cells carrying plasmids encoding the tagged protein of interest and the GFP 1–10 detector fragment (Fig. 1) are mixed with cells expressing red fluorescent protein bearing a 6His tag, plated and grown overnight. *Step 2:* The colony membrane is moved to the induction plate then imaged. Colony fluorescence is proportional to total expression of the protein of interest. Red fluorescent clones will aid in later alignment of the image and colony membrane. *Step 3:* The colony membrane is moved to the capture plate containing metal affinity resin beads (Talon® resin) immobilized in agarose. Colonies are partially lysed by misting with a chemical lysis reagent, releasing the protein of interest fused to fluorescent, complemented GFP. Soluble fluorescent proteins pass through the filtration membrane and bind to the beads. Insoluble proteins remain on membrane. *Step 4:* The colony membrane is returned to the LB agar plate for later picking, and the capture plate is imaged. The Talon®-bound fluorescence reflects soluble protein. *Step 5:* Image processing software is used to align the pictures of the fluorescent colonies and the fluorescent bead capture plate. The images of desired colonies are highlighted, corresponding to brighter spots on the assay plate. *Step 6:* The marked colony membrane image is projected onto the colony membrane using a digital micro-projector, aligned using the red fluorescent clones, and the highlighted clones are picked. (**b**) Application to mock protein library comprised of soluble protein #17 (phosphate cyclase, Table I) in a 25-fold excess of clones expressing insoluble protein #18 (purine-nucleoside phosphorylase, Table I). Image of colonies (*left*), assay plate showing Talon®-bound fluorescent, soluble protein (*middle*), aligned superimposed images (*right*) ($\frac{1}{4}$ s exposure). Clones expressing soluble protein #17, indicated with white arrows. Clones expressing red fluorescent protein aid in the alignment of images.

dilution plates, were washed and plasmids recovered using a QIAprep® kit. Other available ORF filters could also be used (Lutz *et al.*, 2002; Zacchi *et al.*, 2003). Inserts were released by NdeI/BamHI digest (flanking the blunt cloning insertion site and the trapped ORF fragment), sized and gel purified as above, subcloned into the N6His pTET S11 vector and transformed into chemically competent BL21 (DE3), pET GFP 1–10 cells as described (Cabantous *et al.*, 2005). Investigators can expect this protocol to take about 10 days to complete.

Quantification of fluorescence of colonies and Talon® blots for the p85 large fragment library

Bauer plates with $\sim 2 \times 10^3$ colonies were grown overnight at 32°C, then lysed and the bead assay performed as described (Materials and methods, *Colony induction, imaging and partial lysis of clones on capture plate*). A mask was made in Paint ShopPro® and overlaid digitally onto the colony and Talon® blot images so that only the selected targets were visible. The average integrated

fluorescence was tabulated for each colony and corresponding Talon® blots in the masked images using the object analysis function in ImageTool®, exported to Microsoft Excel®, and the ends of the corresponding sequenced ORF fragment used to make the fragment map (Fig. 4a).

Analysis of solubility of selected BCR fragments containing clones in liquid culture

N6His pTET S11 plasmids with target genes for BCR hits identified from the p85 fragment library screen (above) were isolated by retransforming the 400-fold diluted plasmids and selecting only on spectinomycin. A total of 3.5 ml cultures were grown to 0.5 $OD_{600 \text{ nm}}$ as described (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006) (Materials and methods, *Growth of liquid cultures and assay of soluble and insoluble protein in vitro*). Cultures were split into two equal aliquots and cell pellets were frozen overnight at −80°C. An aliquot was thawed, and 500 µl of SoluLyse® was added, mixed and cells were allowed to incubate at 37°C for 1 h prior to
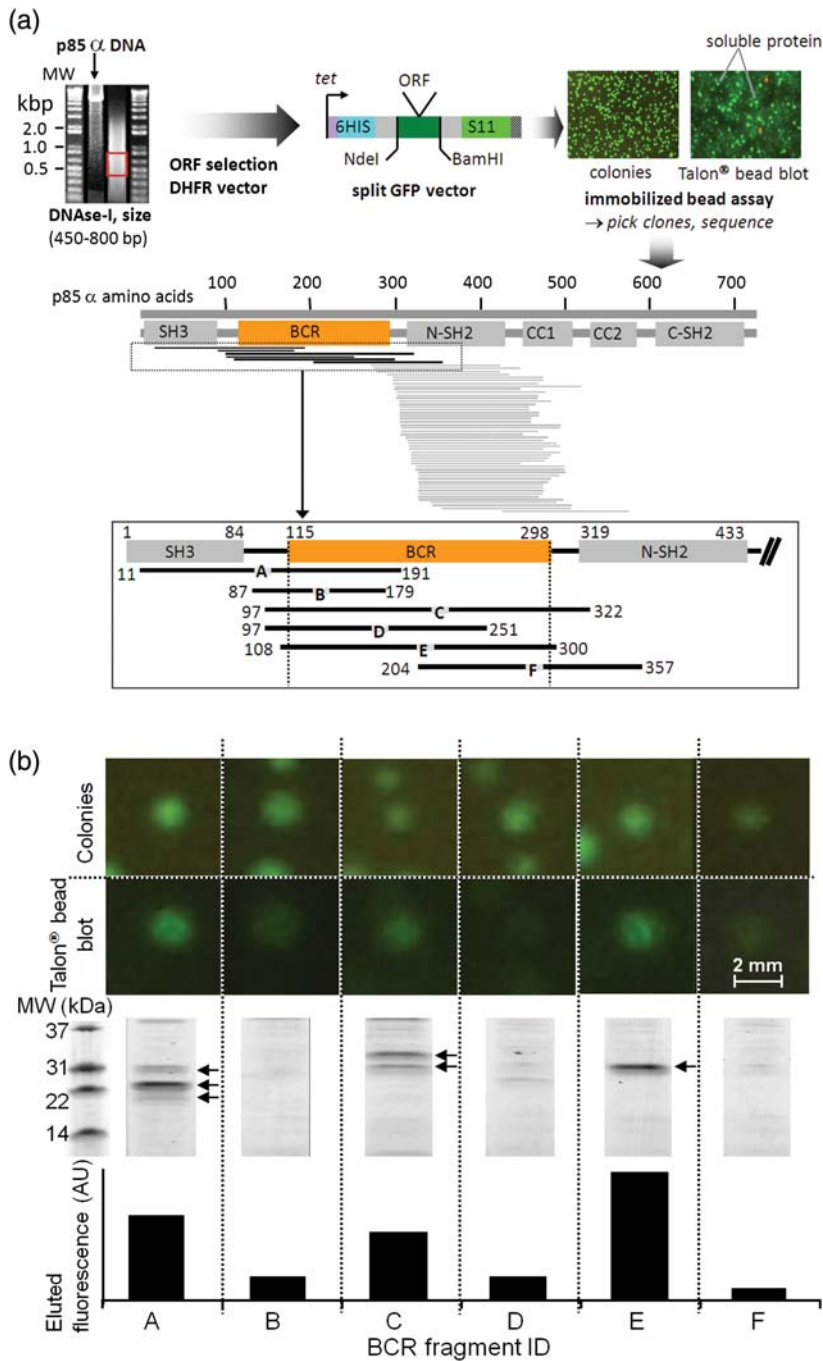
Fig. 4. Results of the p85 large fragment library screen. (**a**) Diagram of experimental scheme (*top*) and of the fragments containing at least half of the BCR domain tagged with GFP S11 (enlarged inset bottom). (**b**) Scale-up expression of the 6 BCR hits. Images of the colonies (from the original Bauer plates of ~4000 colonies per plate) corresponding to the BCR hits ($\frac{1}{4}$ s exposure) (row marked 'colonies') and Talon® capture plate (4 s exposure) (row marked 'Talon® bead blot'). Scanned image of SDS-PAGE gel of Talon®-bound fractions of 3 ml liquid cultures run on a 4–20% SDS-PAGE and stained with Coomassie Blue (*middle row*) to visualize the over-expressed proteins (indicated by arrows). Heavier protein in all the samples around 34 kDa is a contaminating *E.coli* protein. Talon®-bound fluorescence after *in vitro* complementation with GFP 1–10 (*bottom row*).

measurement of soluble and insoluble protein using GFP 1–10 complementation (Cabantous *et al.*, 2005). Eighty microliters of the complemented, soluble fraction was then incubated with 80 resin bed of Talon® resin for 10 min at 22°C, and washed with three aliquots of 200 μl SoluLyse® cocktail, prior to reading on a plate reader (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006). SDS-PAGE and densitometry analyses of soluble, insoluble and Talon®-bound fractions (without the added GFP 1–10) were performed as described (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006).

PCR verification of picked target control proteins and complexes. Sequence analysis of selected p85 library fragments.

Forward primer 5′-TAGAGATACTGAGCACATCAGCAG GACGCACTGACC-3′ and reverse primer 5′GAGGCCT CTAGAGGTTATGCTAGTTATTGC-3′ priming just outside the cloning site of the N6His pTET S11 vector (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006) were used to amplify control proteins and complexes picked from mock libraries. Amplified PCR products were resolved by agarose

gel electrophoresis (Cabantous *et al.*, 2005) with a GelDoc®System (Biorad). p85 library optima were sequenced by fluorescent dye terminator sequencing using the vector-specific primers described above. Sequences were blasted locally against the p85 gene using Bioedit® 7.0.5 (Hall, 1999). The fragment endpoints were analyzed in Microsoft Excel to make the gene maps presented in Fig. 4a.

Results and discussion

Validation of the colony-based 'immobilized bead' assay using control proteins

In the immobilized bead assay depicted in Fig. 1, GFP S11-tagged proteins are coexpressed along with GFP 1–10 to label all the proteins with GFP fluorescence in the cell. Under these conditions, the resulting fluorescence is proportional to the total expressed GFP S11-tagged protein for a wide range of proteins and mutants (Cabantous *et al.*, 2005). Presumably the excess GFP 1–10 present in the cell is available for rapid binding to the GFP S11 tag as soon as it appears off of the ribosome, and before the fused protein has a chance to aggregate. In some cases, for very rapidly aggregating proteins S11 might not be available for binding to GFP 1–10. The colonies are imaged using a digital camera. Next, chemical lysis of some of the cells in each colony releases soluble proteins that diffuse through the Durapore® membrane, through that agarose, and bind to the immobilized Talon® resin via a 6His tag. By misting the chemical lysis over the colonies with a spray bottle, uniform and even lysis of the colonies was easily achieved. The uniform lysis is demonstrated in the 'Talon® bead blot' lane of Fig. 2a where multiple copies expressing the same test proteins show similar bead-bound fluorescence. Imaging the beads gives a measure of the soluble, bead-bound fluorescent protein. Before we could use the assay to report total and soluble Talon®-bound protein we needed to test three aspects that make the assay different from our previous work and that might affect protein solubility: (i) the method of induction, (ii) the lysis chemistry and (iii) the format in which the Talon® beads are used. We also needed to check whether chemical lysis of colonies might lead to cross-contamination of clones during subsequent picking.

The effect of coinduction of S11-tagged proteins and GFP 1–10 on protein solubility

Whole-cell fluorescence is proportional to the expression level of GFP S11-tagged proteins when they are coexpressed with the GFP 1–10 (Cabantous *et al.*, 2005). In the immobilized bead assay, we planned to use the colony fluorescence to monitor the expression level of the test protein. Since the next step in the assay is to lyse the cells and measure the released soluble protein by fluorescence (Supplementary Fig. S3), it is important that the solubility of the labeled protein be as similar as possible to the unlabeled protein. In earlier work, we showed that fusion with intact GFP can substantially reduce protein solubility compared with expressing the proteins alone (Pédelacq *et al.*, 2006). Since coexpression of the GFP 1–10 and the S11-tagged proteins leads to a fused GFP moiety, we wondered whether this might strongly perturb protein solubility. In earlier work we lysed cells expressing GFP S11-tagged proteins, and then added the GFP 1–10 to the clarified lysates to quantify soluble protein (Cabantous *et al.*, 2005). We asked whether GFP S11-tagged proteins behave differently after lysis if they had been pre-complemented in the cell with GFP 1–10. To study the effect of coexpression of GFP 1–10 on the solubility of the GFP S11 proteins in greater detail, we performed split GFP assays *in vitro* on the soluble and pellet fractions of 18 control proteins (Materials and methods) expressed alone or with GFP S11 tags, as described (Cabantous *et al.*, 2005; Cabantous and Waldo, 2006), and tabulated the fraction of protein expressed in soluble form (Table I). In a separate experiment we coexpressed the 18 control proteins with GFP S11 tags along with GFP 1–10 in liquid culture and calculated the fraction of fluorescent protein that was soluble (Materials and methods, Table I). We also measured the Talon®-bound fluorescence for later tests (see below *Assay of total and Talon®-bound protein using the immobilized bead assay; comparison with liquid culture*). The fraction of each protein expressed in soluble form was well correlated for the two methods of expression (linear correlation coefficient $R^2 = 0.88$, Supplementary Fig. S4). We concluded that unlike direct fusions to GFP, labeling the proteins with GFP using the split GFP coexpression protocol did not strongly perturb protein solubility. Likely, this is because the S11 tag had been engineered to not perturb protein solubility, and that during coexpression, the S11-tagged protein has a chance to substantially complete its folding on the ribosome prior to interacting with the GFP 1–10 fragment *in trans*. Furthermore, the GFP 1–10 had also been engineered to not aggregate or complete its folding prior to interacting with the S11 tag, further reducing the likelihood of the GFP 1–10 interfering with the target protein folding. In direct fusions to GFP, the upstream protein can interfere with the folding of the fused GFP domain (Waldo *et al.*, 1999; Pédelacq *et al.*, 2002).

The effect of lysis chemistry on protein solubility of GFP S11-tagged proteins coinduced with GFP 1–10

Sonication is not feasible for the colony-based assay. Earlier we showed that chemical lysis using SoluLyse® gave solubility similar to sonication for a library of acyl carrier protein domain fragments (Listwan *et al.*, 2010). To study the effect of the buffer conditions in SoluLyse® on the solubility of a set of proteins, we compared sonication in SoluLyse® with sonication in TNG buffer for a set of 18 control proteins (Materials and methods). Cultures expressing both GFP 1–10 and proteins tagged with S11 were lysed by sonication in the optimized SoluLyse® cocktail (Materials and methods, *Optimization of lysis cocktail and conditions for 18 controls*) or in TNG, and then the soluble, insoluble and Talon®-bound fluorescence were each measured (Table I). The solubility (soluble protein divided by total protein) of each coinduced sample sonicated in the presence of SoluLyse® was compared with the corresponding solubility obtained using sonication in TNG (Supplementary Fig. S5). Although protein #11 was less soluble when lysed in the presence of SoluLyse® compared with TNG, the two data sets are well correlated (linear correlation coefficient $R^2 = 0.89$). We speculate that protein #11 might be sensitive to the detergent present in SoluLyse®. Alternatively, protein #11 might require one of the components of TNG buffer (for example glycerol) for solubility. We concluded that overall,

the presence of SoluLyse® did not strongly perturb the solubility of the control proteins.

Assay of total and Talon®-bound protein using the immobilized bead assay; comparison with liquid culture

We carried out a set of tests to assess whether our colony-based, immobilized bead assay (Fig. 1) gave results similar to established methods using cell growth in liquid culture followed by sonication and binding of tagged proteins to Talon® beads (Fig. 2a). We photographed the samples of the whole-cell and Talon®-bound fractions from the liquid cultures (Fig. 2a, upper two rows, Table I). To evaluate the immobilized bead assay depicted in Fig. 1, the same coin-duction experiment was carried out using single colonies of the 18 control proteins on Durapore® membranes resting on LB agar (Fig. 2a, row marked 'colonies'). Plotting the fluorescence of the colonies vs. the fluorescence of the liquid culture cell pellet gave a linear correlation coefficient of $R^2 = 0.90$ (Fig. 2b, Table I). Partial lysis of the colonies with SoluLyse® released protein that diffused and bound to the Talon® in agarose. The membrane with the colonies was returned to its original agar plate so the bead plate could be photographed (Materials and methods, *Colony induction, imaging and partial lysis of clones on capture plate*). Fluorescence of the Talon® spots on the capture plate for each control protein (Fig. 2a, lower row marked 'beads in agarose') correlated well ($R^2 = 0.90$) with the fluorescence of the Talon®-bound fraction for the same control protein in liquid culture cell lysates (Fig. 2a, upper row marked 'beads'), giving a linear correlation coefficient $R^2 = 0.90$ (Fig. 2c). It is worth pointing out a potential source of deviations between the two methods for measuring soluble protein (bulk solution vs. immobilized bead assay). Despite centrifugation or filtration through a relatively large-pore filter (i.e. 0.2 µm pore size), soluble aggregates could remain in the soluble fraction and even bind the Talon® column in the bulk solution assay. Such aggregates could be less likely to diffuse through an agarose matrix in the immobilized bead assay. In these cases, the bulk assay would tend to over-estimate the amount of soluble protein. Despite these possibilities, the correlation of the liquid culture experiment and the immobilized bead assay is strong enough to conclude that colony fluorescence is an acceptable surrogate for total protein expression, and that the immobilized bead assay is well correlated with a standard assay measuring protein bound to Talon® resin in batch mode.

Testing for cross-contamination during the chemical lysis of the colonies

Image analysis of colony diameters before and after misting with Solulyze® indicated ∼25% of each colony was lysed. Referring to Fig. 2a the diameter of the spot corresponding to the released protein faithfully reflected the diameter of the colony. Importantly, we found that remaining cell mass was viable—a 96-well plate with nutrient media was filled with picks from colonies after lysis and all wells showed excellent growth similar to colonies that had never been lysed. To test whether cross-contamination might occur due to the colony lysis, we plated a 50 of 50 mix of colonies expressing either red fluorescent or green fluorescent proteins at high density (∼3000 colonies on a Bauer plate). Red and green colonies were picked after lysis and streaked out. No red clones were found in green streaks and vice versa, indicating that there was no cross-contamination (results not shown). We conclude that clones can be recovered without the need for replica plates, a significant advantage for high-throughput protein screens.

Picking of mock binary libraries guided by a digital projector; application to protein solubility screen

We tested how the images of the colonies and the bead-bound fluorescence could be used to guide colony picking. As shown in Fig. 3a, a mock library was made by mixing *E.coli* stocks expressing soluble control protein #17 and insoluble control protein #18 (Table I) in a 1 : 25 ratio. These control proteins were selected because they are similar in expression level (see Fig. 3b, left), but differ greatly in solubility. To facilitate the alignment of the capture plate image with the plate containing the remaining viable colony mass, the library was spiked with *E.coli* expressing soluble, 6His-tagged DsRED protein. Colonies expressing the soluble control protein #17 yielded fluorescent, bead-bound spots on the capture plate (indicated by arrows in Fig. 3b, middle). The capture plate image was projected onto the plate containing the partially lysed colonies (Fig. 3b), and aligned using the DsRED fluorescent colonies (Fig. 3a, Step 6) as depicted by the overlay of the colony and Talon® images (Fig. 3b, right). Several clones corresponding to soluble spots were picked using the guide marks. All grew in selective media and sequencing confirmed they corresponded to control protein #17 as expected.

Application of the colony-based 'immobilized bead' assay to screening an ORF fragment library for soluble proteins

We tested the utility of our colony-based assay by applying it to the problem of identifying soluble modules of a large protein. Referring to Fig. 4a, the method was used to screen a cDNA fragment library of the *H.sapiens* p85 protein, a large, multi-domain protein, for soluble protein constructs containing the breakpoint cluster region-homology (BCR) domain. To increase the chances of selecting hits for the large BCR domain, we fragmented the entire p85 gene by DNAse-I and size selected DNA fragments between 400 and 800 bp long (Fig. 4a, Methods, *Making the p85 large fragment library, ORF selection and cloning*). To select for in-frame fragments prior to cloning the library into the 6His-X-S11 vector, the fragments were cloned into an internal permissive site of DHFR (Fig. 4a). Because the translations of both halves of the DHFR are required for antibiotic selection, and the fragments must be linked (in this case by the trapped cDNA ORF fragment), the method should select against internal stop codons (enriching in in-frame fragments) and also select against *de novo* internal ribosome-binding sites. Further details and validation of this ORF filter will be described in a subsequent manuscript (Pédelacq *et al.*, under review). A less stringent but more traditional fusion to C-terminal DHFR (Dyson *et al.*, 2008) or other available ORF filters that more stringently select against internal ribosome entry sites as does our 'insertion' DHFR could have been used to provide the selection (Lutz *et al.*, 2002; Zacchi *et al.*, 2003). Transformed clones were then grown in the presence of TMP to select for in-frame fragments. The resulting library was >10⁶ colony-forming units with maximum diversity coverage. Further details and

validation of this ORF filter will be published in a subsequent manuscript (Pédelacq *et al.*, under peer review).

Making the ORF selection library is the most laborious step of this method, taking an investigator ∼10 days to complete. For small screening campaigns, the investigator can choose to omit the ORF enrichment step, and simply screen more clones to compensate accordingly. Subsequent steps can be accomplished in a couple of days. Pooled inserts of survivors were then sub-cloned *en masse* into the S11 tagging vector in preparation for the immobilized bead assay. A sample of the library was spiked with clones expressing DsRED to facilitate alignment of the colony plate and capture plate images for analysis and optima selection. Approximately 8000 clones from the ORF selection library were screened. After photographing the colonies and the Talon® resin plate, we used PaintShop Pro® (JASC) to align the images in a stack for analysis based on the position of the randomly placed DsRED clones. We chose 96 colonies, preferentially picking those with Talon®-bound fluorescence, but also including some with little or no soluble fluorescence. Clones showing no colony fluorescence were avoided. Sequencing of these 96 colonies revealed most contained the central NSH2 domain and some flanking regions (Fig. 4a), and a few contained part or the entire BCR domain. Many of the NSH2-containing constructs appeared to be better expressed than the other domains (data not shown). We suspect this increased expression level may have led to enrichment for NSH2 during the ORF selection (Materials and methods, *p85 large fragment library, ORF selection and cloning*). Since we had deliberately not included shorter fragments near the expected size of the NSH2, most picks containing NSH2 were longer than the hypothetical boundaries. Because our objective focused on the BCR domain, these NSH2 clones were not pursued further. Six clones (Fig. 4a–f, Supplementary Fig. S6) contained at least half of the BCR domain, and two of these contained the full-length BCR domain (Fig. 4a, c and E, Supplementary Fig. S6). Clone E is the most compact construct that also contains the entire BCR domain.

To test how well the immobilized bead assay predicted success in expression of soluble protein in liquid culture, plasmids for the BCR fragments (Clones A–F) were isolated and retransformed without the GFP 1–10 (Materials and methods, *Analysis of solubility of selected BCR fragments containing clones in liquid culture*), expressed and the Talon®-bound fractions visualized by SDS gel electrophoresis (Cabantous *et al.*, 2005) (Fig. 4b middle and Fig. 4b legend). To more precisely measure the soluble BCR protein in each sample, refolded GFP 1–10 was added to the soluble lysate (Cabantous *et al.*, 2005), and then the fluorescent complemented proteins were bound to Talon®. The fluorescence was measured on a plate reader (Fig. 4b, bottom). The amount of protein as indicated by the Talon®-bound fluorescence of the immobilized bead assay (Fig. 4b, top) was well correlated with the liquid culture Talon®-bound protein as visualized by SDS-PAGE (Fig. 4b, middle) and *in vitro* complementation fluorescence (Fig. 4b, *bottom*). Clones B, D and F were each very faint in the colony-based immobilized bead assay, and are not visible on SDS-PAGE as expected (Fig. 4b). Clones A, C and E showed up brightly on the immobilized bead assay, were readily visible on

SDS-PAGE and complemented GFP 1–10 well (Fig. 4b, bottom).

We note that the sequences for BCR domain Clones A and F each contained an extra base insertion at the 5′ cloning site that resulted in predicted frame shifts and stop codons in the frames of the fragments (Supplementary Figs S7 and S8). Both clones passed the ORF selection step (Methods and materials *p85 large fragment library, ORF selection and cloning*) and had detectable Talon® bead-bound fluorescence (Fig. 4b) implying that translation must continue beyond the stop codon and that both ends of the polypeptide are covalently linked. Such artifacts can result from ribosome frameshift and reinitiation (Adhin and van Duin, 1990) and have been reported elsewhere (Goldman *et al.*, 2000). See the *Supplementary discussion* for a more complete analysis and discussion of the frame-shift artifacts. Clone C produced two bands on SDS-PAGE, perhaps because it includes additional 'linker' sequence on the ends and might be partially proteolyzed after work-up (Graslund *et al.*, 2008a,b). On the other hand, Clone E produced a single, intense band at ∼27.7 kDa, close to the expected molecular weight of 27.1 kDa. It has very little 'additional' linker sequence on its ends and is the most compact, soluble domain of the set (Fig. 4, Supplementary Fig. S6). Importantly, the ends of fragment E (amino acids 108–300) correspond closely to the construct from which the structure was determined (Musacchio *et al.*, 1996) (amino acids 105–319). Removing unstructured regions of proteins by engineering the ends of the coding sequence or by proteolysis can improve soluble expression and crystallizability (Gao *et al.*, 2005; Hart and Tarendeau, 2006; Angelini *et al.*, 2009).

Application to finding multi-protein complexes amenable to co-purification; validation using eight control complexes

Given the biological importance of protein assemblies, there is increasing interest in solving the structures of multi-protein complexes (Derewenda and Vekilov, 2006; Strong *et al.*, 2006; Garcia *et al.*, 2009). We developed a variation of our bead-based assay that corresponds to the widely used 'tandem affinity purification' strategy (Graslund *et al.*, 2008a) and used it to confirm protein complex assembly by tagging one member of a complex with an affinity tag, and the other with GFP S11. In all, eight controls (Table II) were cloned by PCR as polycistrons from genomic DNA (Supplementary Table S1). Stable trimeric and dimeric complexes were tested. We also studied alternative versions of the constructs where one subunit was omitted to make them dissociate (Table II). For each multi-protein complex, the first subunit encoded at the front of each polycistron was tagged with the N-terminal 6His tag and the last subunit with the C-terminal GFP S11 tag (Fig. 5a). From this point forward, we followed the same basic procedure as described above (Materials and methods). The eight constructs were coinduced with the GFP 1–10 detector fragment to label the S11-tagged subunit with fluorescence (Fig. 5b, top). After lysis, we expected that all soluble subunits of the complexes would travel through the Durapore® membrane and enter the agarose, regardless of their state of assembly. Stable assembled complexes would bind Talon® beads via the N-terminal 6His tag, forming distinct fluorescent dots under
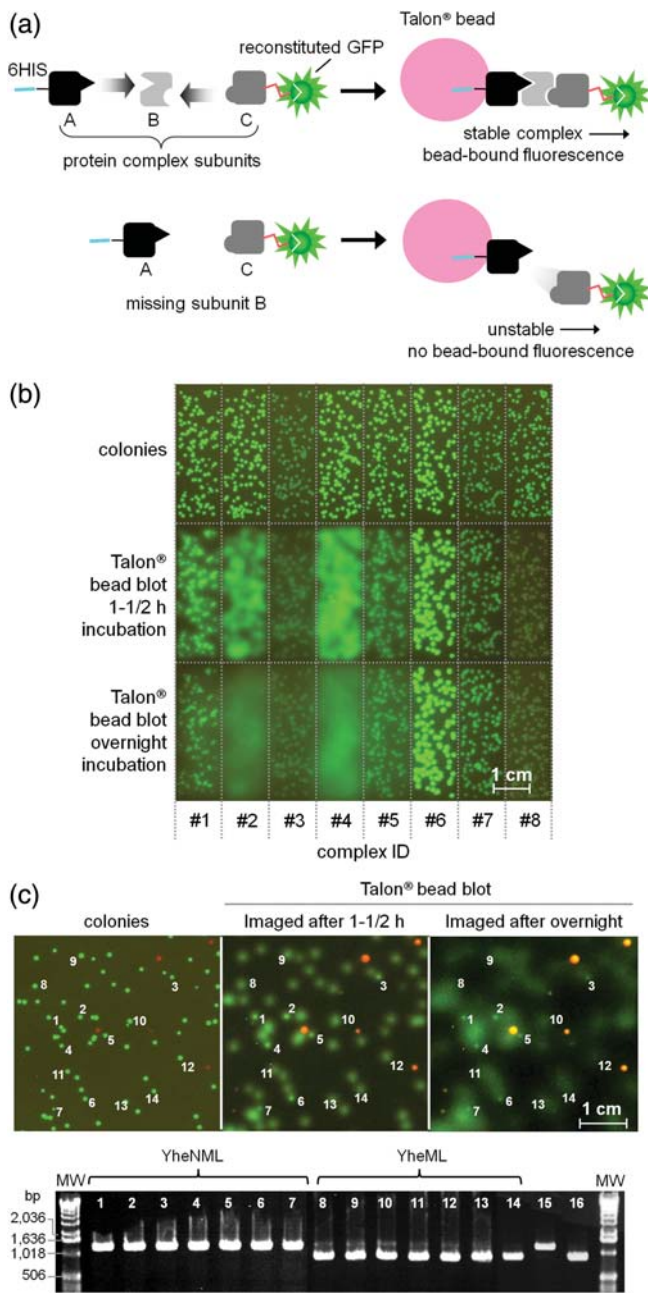
the colonies. On continued incubation, the unassembled complex subunits would diffuse away from the Talon® beads. Photographs were taken after $1-\frac{1}{2}$ h at room temperature (Fig. 5b, middle), then after 18 h incubation in at 4°C (Fig. 5b bottom) to observe this diffusion.

In agreement with these expectations (Fig. 5b), Talon®-bound fluorescence was observed for all the complexes for which structures had been previously determined. These included the *E.coli* trimer YheNML (Numata *et al.*, 2006) (Fig. 5b, #1), the *Mycobacterium tuberculosis* heterodimer Rv2431c/Rv2430c (Strong *et al.*, 2006) (Fig. 5b, #3) and the heterodimer allophanate hydrolase from *M.smegmatis* (Fig. 5b, #8) for which a structure was recently solved by Eisenberg and coworkers (M. Kaufmann, personal communication, DOI:10.2210/pdb3mml/pdb). Three close homologs of the hydrolase (Fig. 5b, control complexes #5–7) also bound beads. Conversely, only diffuse fluorescence was observed when a member of the complex was omitted (Fig. 5b, #2 and #4). As depicted in Fig. 5b, the omission of the YheN subunit from the trimer YheNML (Fig. 5b, #2) is hypothesized to destabilize the complex formation, no longer coupling the fluorescent subunit with the bound subunit. The omission of Rv2431c from the dimer Rv2431c/Rv2430c (Fig. 5b, #4) may result in an inaccessible 6His tag, hence the lack of Talon®-bound fluorescence.

Identification of soluble, stable complexes in mock libraries

We carried out a test to determine whether it was feasible to use the bead-based assay to isolate soluble, stable complexes. We combined clones of YheNML with YheML in a 1:25 ratio to make a simple binary library. Following the protocol for the control complexes, colonies (Fig. 5c, top) were coinduced (Fig. 5c, left), lysed, the capture plates allowed to incubate and photographs were taken at $1-\frac{1}{2}$ h (Fig. 5c, middle) and after overnight incubation (Fig. 5c, right) to allow further diffusion of unassembled complexes. The overnight capture plate image was projected and aligned onto the plate containing the colonies. Talon®-bound fluorescence identified colonies expressing the assembled complex. Several clones were picked and PCR screens showed that the colonies with compact Talon®-bound fluorescence in the bead blot corresponded to YheNML clones (Fig. 5c, bottom).

Conclusion

The immobilized bead assay is a simple assay for protein expression and solubility. It maintains colony viability, avoids replica plates and uses inexpensive, readily available materials and equipment. Talon®-bound soluble protein fluorescence in the immobilized bead assay, as measured using an inexpensive digital camera, agrees closely with bound fluorescence measured for beads in a micro-well format using a plate reader (Fig. 2c). Proteins and protein complexes are directionally immobilized on Talon® under conditions that closely resemble traditional batch or column affinity purification (Fig. 1) (Graslund *et al.*, 2008a). This makes it possible to distinguish when the subunits of protein complexes are assembled or merely soluble but dissociated (Fig. 5). The example of YheML (missing the YheN subunit) shows that the assay can be useful even in the presence of significant background. The tagged YheL subunit remains soluble, adding to the background fluorescence on the plate.

Fig. 5. Application of GFP colony filtration immobilized bead assay to detect assembled protein complexes. (**a**) Principle of bead-binding assay for complexes. A multi-protein complex carries a 6His tag on one subunit, and the GFP S11 tag on another subunit for labeling with GFP fluorescence (*top*). After lysis and filtration for soluble proteins, bead-bound fluorescence indicates intact assembled complexes (*top*). Lack of Talon®-bound fluorescence indicates insoluble or unstable complexes, for example, those missing a required subunit (*bottom*). (**b**) Validation of the GFP colony filtration bead assay using eight control multi-protein complexes (Table II) tagged as in (a). Image of colonies after coexpression ($\frac{1}{5}$ s exposure) (*top*). Talon resin capture plate $1-\frac{1}{2}$ h after partial colony lysis and removal of colony membrane (*middle*), and after subsequent overnight incubation at 10°C (4 s exposure) (*bottom*). Compact fluorescent spots indicate assembled soluble complexes. (**c**) Application to a library consisting of cells expressing stable YheNML or unstable YheML (b columns 1 and 2). Image of colonies ($\frac{1}{4}$ s exposure) (*left*), capture plate $1-\frac{1}{2}$ h after lysis and transfer of colony membrane (*middle*), capture plate after additional overnight incubation at 4°C (4 s exposure) (*right*). Image of PCR products amplified from DNA of picked clones (*bottom*). Picks with compact Talon® blots (colonies marked 1–7) give PCR products (Lanes 1–7) with the same mass as YheNML control (Lane 15). Picks from clones with diffuse Talon® blots (colonies marked 8–14) match the YheML control (Lane 16). Mass standards 1 kb Plus (Invitrogen) (lanes marked MW).

Nonetheless, the stably assembled YheNML complexes are easily identified by compact Talon® blot spots, even as early as $1-\frac{1}{2}$ after lysis and binding (Fig. 5). Such clear identification is not possible when all of the subunits are non-specifically captured on nitrocellulose regardless of their state of assembly (Cornvik et al., 2005, 2006; Dahlroth et al., 2006).

Beads are to be expected to have a higher binding capacity than membranes due to their larger surface area, facilitating direct visualization of fluorescently labeled proteins. Other beads could in principle be used to capture the protein (Graslund et al., 2008a). For example, amylose could be used to capture maltose-binding protein tags (Nallamsetty and Waugh, 2006, 2007), or glutathione-conjugated beads to capture glutathione-S-transferase tags (Goda et al., 2004), provided that the agarose was cooled prior to adding the resin. However, Talon® resin is widely used, and we have found that it is sufficiently stable to be recovered after the bead assay by melting the agarose in a water bath and allowing the Talon® beads to settle down (G.S.Waldo, unpublished).

We use coinduction of the split GFP fragments to label all of the expressed protein with fluorescence. This enables an estimate of total protein expression using total colony fluorescence, and soluble protein by examining Talon®-bound fluorescence (Fig. 2a) in one experiment. However, it should be pointed out that the proteins could be labeled in other ways. For example, protein stain reagents such as SyPro Orange (Invitrogen) could be added to the Talon® resin plate after capture of the released proteins (Fig. 3a, Step 3). Proteins could be tagged with the FLAG tag then detected with a labeled anti-FLAG antibody (Perrin et al., 2001; Shukla et al., 2007), although the Talon® resin plates would have to be washed to remove unbound fluorescent probes.

We have also used GFP 1–10 as a reagent to label the Talon® resin-captured proteins (G.S.Waldo, unpublished). One advantage of the GFP 1–10 as an in vitro assay reagent is that it only becomes fluorescent when it binds to the GFP S11 tag, eliminating the need to wash away excess GFP 1–10 prior to visualizing the complemented protein. Joly and coworkers recently used purified GFP 1–10 as a protein reagent to visualize S11-tagged proteins in fixed and permeabilized mammalian cells (Kaddoum et al., 2010). The authors reported that the split GFP labeling had a much lower background compared to protocols detecting the targets with conventional labeled antibodies. Other experimental parameters include differential effects on protein solubility arising from the labeling method using coexpression of split GFP and the chemical composition of lysis buffers, non-uniformity of illumination during photography of beads and colonies, and non-linearity caused by the dynamic range of the camera used to photograph the beads. These variables might need to be optimized for different classes of proteins. For example, the lysis cocktail might need additional adjuvants or cofactors to ensure protein stability for specific applications. Alternative lysis methods could be used such as freeze−thaw of the colony membrane (Cornvik et al., 2005), prior to transfer of the membrane to the Talon® capture plate.

Recent work suggests that compact versions of protein domains (or combinations of domains) are more likely to crystallize and diffract well compared with proteins that contain disordered elements such as N- or C-terminal extensions (Gao et al., 2005). It should be pointed out that although we found a fragment of p85 that expressed relatively compact truncation variants of the BCR and NSH2 domains, we did not screen enough fragments to precisely map out the domain boundaries. Rich sampling of ORF libraries for annotating protein domain boundaries and identification of stable protein complexes is the subject of ongoing research in our labs (Pédelacq, Waldo, Terwilliger and co-workers). We envision that the immobilized bead assay described here will be applied to a broad scope of biological problems in addition to protein crystallization target selection and the identification of stably assembled complexes. One such application that has yet to be explored is host strain engineering, where the increased functional or soluble expression of target host proteins can be used to select for chromosomal mutations that improve protein production (Belin et al., 2004; Massey-Gendel et al., 2009).

Supplementary data

Supplementary data are available at *PEDS* online.

References

Adhin,M.R. and van Duin,J. (1990) *J. Mol. Biol.*, **213**, 811–818.
Angelini,A., Tosi,T., Mas,P., Acajjaoui,S., Zanotti,G., Terradot,L. and Hart,D.J. (2009) *FEBS J.*, **276**, 816–824.
Armstrong,N., de Lencastre,A. and Gouaux,E. (1999) *Protein. Sci.*, **8**, 1475–1483.
Belin,P., Dassa,J., Drevet,P., Lajeunesse,E., Savatier,A., Boulain,J.C. and Menez,A. (2004) *Protein Eng. Des. Sel.*, **17**, 491–500.
Cabantous,S. and Waldo,G.S. (2006) *Nat. Methods*, **3**, 845–854.
Cabantous,S., Terwilliger,T.C. and Waldo,G.S. (2005) *Nat. Biotechnol.*, **23**, 102–107.
Cabantous,S., Rogers,Y., Terwilliger,T.C. and Waldo,G.S. (2008) *PLoS One*, **3**, e2387.
Cornvik,T., Dahlroth,S.L., Magnusdottir,A., Herman,M.D., Knaust,R., Ekberg,M. and Nordlund,P. (2005) *Nat. Methods*, **2**, 507–509.
Cornvik,T., Dahlroth,S.L., Magnusdottir,A., et al. (2006) *Proteins*, **65**, 266–273.
Dahlroth,S.L., Nordlund,P. and Cornvik,T. (2006) *Nat. Protoc.*, **1**, 253–258.
Derewenda,Z.S. and Vekilov,P.G. (2006) *Acta. Crystallogr. D Biol. Crystallogr.*, **62**(Pt 1), 116–124.
Dyson,M.R., Perera,R.L., Shadbolt,S.P., Biderman,L., Bromek,K., Murzina,N.V. and McCafferty,J. (2008) *Nucleic Acids Res.*, **36**, e51.
Fujiwara,K., Ishihama,Y., Nakahigashi,K., Soga,T. and Taguchi,H. (2010) *EMBO J.*, **29**, 1552–1564.
Gao,X., Bain,K., Bonanno,J.B., et al. (2005) *J. Struct. Funct. Genomics*, **6**, 129–134.
Garcia,B.L., Tzekou,A., Ramyar,K.X., McWhorter,W.J., Ricklin,D., Lambris,J.D. and Geisbrecht,B.V. (2009) *Acta. Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **65**(Pt 5), 482–485.

Goda,N., Tenno,T., Takasu,H., Hiroaki,H. and Shirakawa,M. (2004) *Protein. Sci.*, **13**, 652–658.

Goldman,E., Korus,M. and Mandecki,W. (2000) *FASEB J.*, **14**, 603–611.

Graslund,S., Nordlund,P., Weigelt,J., *et al*. (2008a) *Nat. Methods*, **5**, 135–146.

Graslund,S., Sagemark,J., Berglund,H., *et al*. (2008b) *Protein. Expr. Purif.*, **58**, 210–221.

Hall,T.A. (1999) *Nucleic Acids. Symp. Ser* **41**, 95–98.

Hart,D.J. and Tarendeau,F. (2006) *Acta Crystallogr. D Biol. Crystallogr.*, **62**(Pt 1), 19–26.

Jacobs,S.A., Podell,E.R., Wuttke,D.S. and Cech,T.R. (2005) *Protein Sci.*, **14**, 2051–2058.

Jacobs,S.A., Podell,E.R. and Cech,T.R. (2006) *Nat. Struct. Mol. Biol.*, **13**, 218–225.

Kaddoum,L., Magdeleine,E., Waldo,G.S., Joly,E. and Cabantous,S. (2010) *Biotechniques*, **49**, 727–728, 730, 732 passim.

Kawasaki,M. and Inagaki,F. (2001) *Biochem. Biophys. Res. Commun.*, **280**, 842–844.

Knaust,R.K. and Nordlund,P. (2001) *Anal. Biochem.*, **297**, 79–85.

Listwan,P., Terwilliger,T.C. and Waldo,G.S. (2009) *J. Struct. Funct. Genomics*, **10**, 47–55.

Listwan,P., Pédelacq,J.D., Lockard,M., Bell,C., Terwilliger,T.C. and Waldo,G.S. (2010) *J. Struct. Funct. Genomics*, **11**, 41–49.

Lutz,S., Fast,W. and Benkovic,S.J. (2002) *Protein Eng.*, **15**, 1025–1030.

Massey-Gendel,E., Zhao,A., Boulting,G., *et al*. (2009) *Protein Sci.*, **18**, 372–383.

Musacchio,A., Cantley,L. and Harrison,S. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 14373–14378.

Nallamsetty,S. and Waugh,D. (2006) *Protein Expr. Purif.*, **45**, 175–182.

Nallamsetty,S. and Waugh,D.S. (2007) *Nat. Protoc.*, **2**, 383–391.

Numata,T., Fukai,S., Ikeuchi,Y., Suzuki,T. and Nureki,O. (2006) *Structure*, **14**, 357–366.

Pédelacq,J.D., Piltch,E., Liong,E.C., *et al*. (2002) *Nat. Biotechnol.*, **20**, 927–932.

Pédelacq,J.D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) *Nat. Biotechnol.*, **24**, 79–88.

Perrin,M.H., Fischer,W.H., Kunitake,K.S., *et al*. (2001) *J. Biol. Chem.*, **276**, 31528–31534.

Puig,O., Caspary,F., Rigaut,G., *et al*. (2001) *Methods*, **24**, 218–229.

Reich,S., Puckey,L.H., Cheetham,C.L., *et al*. (2006) *Protein Sci.*, **15**, 2356–2365.

Rigaut,G., Shevchenko,A., Rutz,B., Wilm,M., Mann,M. and Seraphin,B. (1999) *Nat. Biotechnol.*, **17**, 1030–1032.

Shukla,A.K., Haase,W., Reinhart,C. and Michel,H. (2007) *Int. J. Biochem. Cell Biol.*, **39**, 931–942.

Strong,M., Sawaya,M.R., Wang,S., Phillips,M., Cascio,D. and Eisenberg,D. (2006) *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.

Tarendeau,F., Boudet,J., Guilligay,D., *et al*. (2007) *Nat. Struct. Mol. Biol.*, **14**, 229–233.

Tarendeau,F., Crepin,T., Guilligay,D., Ruigrok,R.W., Cusack,S. and Hart,D.J. (2008) *PLoS Pathog.*, **4**, e1000136.

Terwilliger,T.C. (2004) *Nat. Struct. Mol. Biol.*, **11**, 296–297.

Terwilliger,T.C., Stuart,D. and Yokoyama,S. (2009) *Annu. Rev. Biophys.*, **38**, 371–383.

Waldo,G.S. (2003a) *Curr. Opin. Chem. Biol.*, **7**, 33–38.

Waldo,G.S. (2003b) *Methods Mol. Biol.*, **230**, 343–359.

Waldo,G.S., Standish,B.M., Berendzen,J. and Terwilliger,T.C. (1999) *Nat. Biotechnol.*, **17**, 691–695.

Wang,H. and Chong,S. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 478–483.

Yumerefendi,H., Tarendeau,F., Mas,P.J. and Hart,D.J. (2010) *J. Struct. Biol.*, **172**, 66–74.

Zacchi,P., Sblattero,D., Florian,F., Marzari,R. and Bradbury,A.R. (2003) *Genome Res.*, **13**, 980–990.