

# Automated, high-throughput platform for protein solubility screening using a split-GFP system

Pawel Listwan · Thomas C. Terwilliger ·  
Geoffrey S. Waldo

Received: 7 August 2008 / Accepted: 6 November 2008 / Published online: 28 November 2008  
© The Author(s) 2008. This article is published with open access at [Springerlink.com](http://Springerlink.com)

**Abstract** Overproduction of soluble and stable proteins for functional and structural studies is a major bottleneck for structural genomics programs and traditional biochemistry laboratories. Many high-payoff proteins that are important in various biological processes are “difficult to handle” as protein reagents in their native form. We have recently made several advances in enabling biochemical technologies for improving protein stability (<http://www.lanl.gov/projects/gfp/>), allowing stratagems for efficient protein domain trapping, solubility-improving mutations, and finding protein folding partners. In particular split-GFP protein tags are a very powerful tool for detection of stable protein domains. Soluble, stable proteins tagged with the 15 amino acid GFP fragment (amino acids 216–228) can be detected *in vivo* and *in vitro* using the engineered GFP 1–10 “detector” fragment (amino acids 1–215). If the small tag is accessible, the detector fragment spontaneously binds resulting in fluorescence. Here, we describe our current and on-going efforts to move this process from the bench (manual sample manipulation) to an automated, high-throughput, liquid-handling platform. We discuss optimization and validation of bacterial culture growth, lysis protocols, protein extraction, and assays of soluble and insoluble protein in multiple 96 well plate format. The optimized liquid-handling protocol can be used for rapid determination of the optimal, compact domains from single ORFs, collections of ORFs, or cDNA libraries.

**Keywords** Automation · High-throughput screening · Green fluorescent protein · Protein expression · Protein solubility · Robotics

## Abbreviations

GFP Green fluorescent protein  
ORF Open reading frame  
SAMI SAGIAN<sup>TM</sup> automated method interface

## Introduction

The success of structural genomics programs around the world relies on the supply of soluble protein targets for structure determination. For this reason numerous protein solubility screens have previously been established including screens based on antibody blots [1–3], split  $\beta$ -galactosidase [4, 5] fluorogenic biarsenical FLaSH or ReASH substrates [6], green fluorescent protein (GFP) fusions [7, 8] and others. The development of split-GFP system in our laboratory using optimized derivatives of “superfolder” GFP allows tagging and detection of both soluble and insoluble protein targets *in vivo* and *in vitro*, without perturbing protein folding [9–12]. The development of the method and detailed protocols using standard laboratory bench techniques have been previously described [10]. Briefly, the method uses engineered self complementing GFP fragments derived from “super folder” GFP, a 15 amino acid GFP fragment—strand 11 (S11 or GFP 11) and a separately expressed GFP 1–10 “detector” fragment. The S11 fragment is expressed as C-terminal fusion with the protein of interest in a pTET plasmid and GFP 1–10 is separately expressed from a pET plasmid. If the protein of interest is stably expressed and

P. Listwan · T. C. Terwilliger · G. S. Waldo (✉)  
Bioscience Division, MS-M888, Los Alamos National  
Laboratory, Bikini Atoll Rd, SM30, Los Alamos,  
NM 87545, USA  
e-mail: [waldo@lanl.gov](mailto:waldo@lanl.gov)

soluble, the S11 fragment of GFP is available for complementation by the independently expressed GFP 1–10 fragment, leading to formation of the fluorescent GFP  $\beta$ -barrel. Should the expression and folding of the target protein be compromised by the nature of the protein itself, expression conditions, lysis or purification, the resulting fluorescence will be reflective of the proportion of the soluble fraction that is retained. This powerful method allows screening for protein solubility both *in vivo* and *in vitro*. It can be used to screen collections of ORFs of interest, complete or focused cDNA libraries, or domain trapping. The soluble constructs that are identified are suitable for protein purification, crystallization and many other downstream applications.

The original assay using split-GFP was developed using standard molecular biology methodologies and required manual preparation and execution of assays. Although in many steps multi-channel pipettors can be used, screening of hundreds of candidate constructs proved to be very laborious. Moving from the laboratory bench to automated, high-throughput processing of the clones selected from *in vivo* screening for the higher accuracy *in vitro* assay allows greater processing power of target proteins, reduced human error and elimination of tedious time and labor-intensive handling of single protein targets. The domain trapping and solubility screening using split-GFP involves integration of a number of individual steps, from selection of targets, gene fragmentation, ORF selection, identification of soluble clones, sequencing, and mapping. The most laborious step by far is the *in vitro* screening of hundreds of clones selected from *in vivo* screening. Here, we describe the steps involved in moving the *in vitro* solubility screen assay from the laboratory bench to an automated, high-throughput processing using robotics. The process involves optimizing each individual step and identifying critical points, to produce results comparable to manual manipulation for a number of control proteins.

### Materials, robotics hardware components and software applications

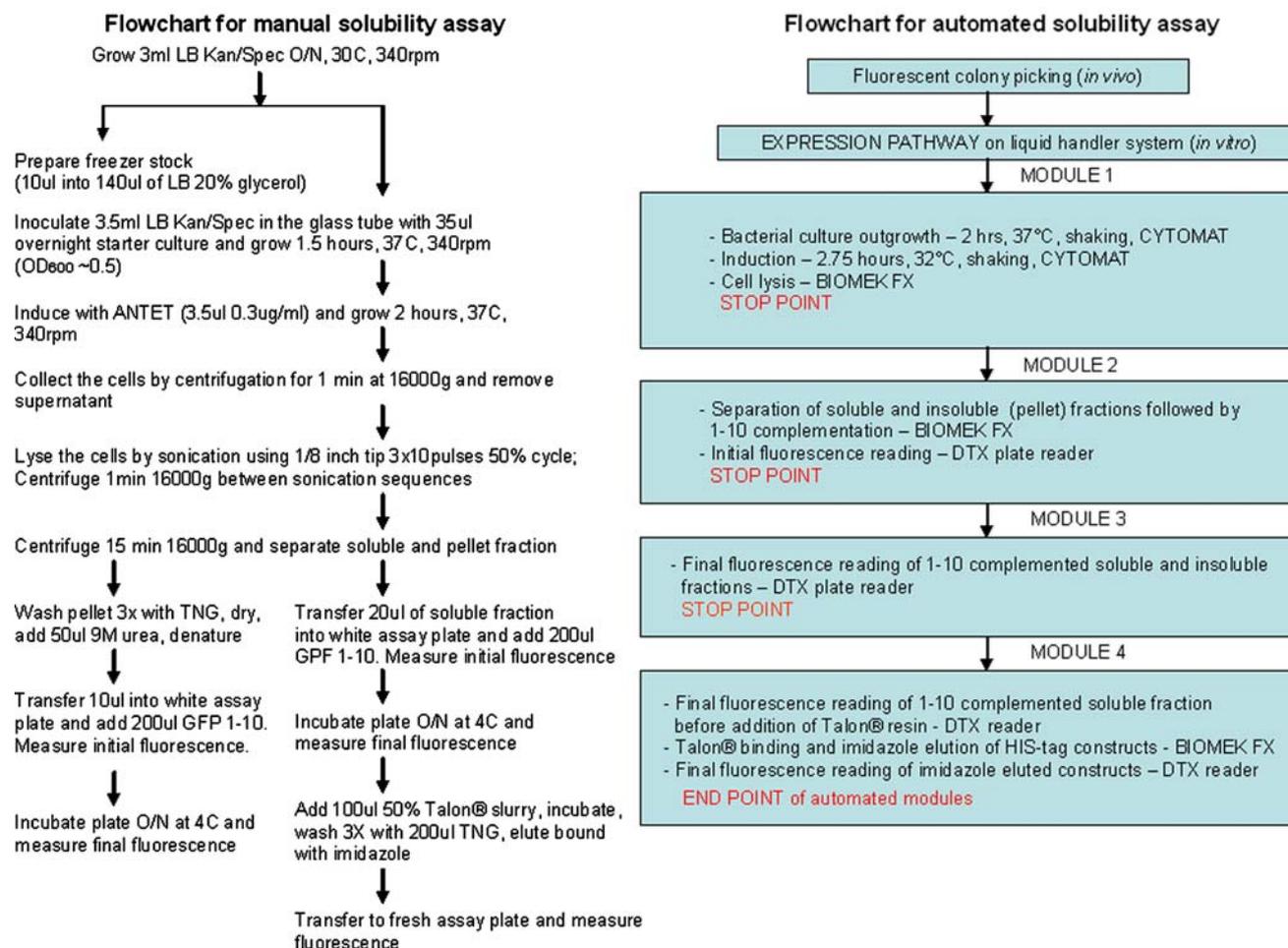
Our integrated high-throughput robotic system used for solubility screening includes a Biomek FX liquid-handling robot, an ORCA<sup>®</sup> arm, a DTX plate reader equipped with filters allowing measurement of both absorbance and fluorescence (Beckman–Coulter, Fullerton, CA), a Cytomat 24 Hotel, Cytomat 2C incubators (ThermoFisher Scientific, Waltham, MA) and a Rotanta 46 ESC centrifuge (Hettich AG, Tuttlingen, Germany). For fluorescence imaging we use an Illumatool lighting system LT-9500 (Lighttools Research, Encinitas, CA). Overnight culture

growth is performed in Innova 4230 refrigerated incubator shakers (New Brunswick Scientific, Edison, NJ). Manual sonication of the 96-well plates is performed using a Sonicator—ultrasonic processor XL20-20 (Misonix Inc., Farmingdale, NY). All Biomek FX methods are built using Biomek Software, version 3.2. Integration of all robotics components is controlled by SAMI Method Editor, version 3.5 (Beckman–Coulter, Fullerton, CA).

All plasmids are expressed in *E. coli* BL21 (DE3) strain (Stratagene, La Jolla, CA). For overnight growth we use Luria–Bertani (LB) media supplemented with 7.5% glycerol and selective antibiotics. Protein expression is induced by Anhydrotetracycline (ANTET) and quenched using Chloramphenicol (CAT) (Sigma–Aldrich, St. Louis, MO). Buffers used in all steps include either TNG buffer (100 mM Tris–HCl (pH 7.4), 150 mM NaCl, 10% glycerol) or TN buffer (same as TNG but without glycerol). Talon<sup>®</sup> Superflow metal affinity resin (Clontech, Mountain View, CA) is used for binding 6-His tag containing soluble proteins. Elution of the 6-His tag proteins is done with 250 mM (final concentration) imidazole (Sigma–Aldrich, St. Louis, MO) in TN buffer. Chemical cell lysis was performed using lysozyme (Sigma–Aldrich, St. Louis, MO), SoluLyse (Genlantis, San Diego, CA) or Bugbuster protein extraction reagent (Novagen, EMD Chemicals Inc., San Diego, CA). All chemical lysis reagents were supplemented with Benzonase (Sigma–Aldrich, St. Louis, MO). GFP 1–10 complementation fragment was prepared as previously described [10] and resuspended in TNG buffer.

### Manual versus automated *in vitro* solubility screening using split-GFP

Depending on the number of the protein targets to be processed it is a choice for the researcher to use either manual or automated assays for assessment of protein solubility. Figure 1 represents a flowchart of the *in vitro* solubility screen manual method side by side with the automated method. The flowchart includes individual steps from overnight culture growth, inoculation of the overnight starter culture, cell lysis, soluble and insoluble fraction separation, to complementation of both soluble and insoluble fractions. Adaptation of the manual method for an automated, high-throughput sample processing platform required modification of several steps and adjustment of several experimental parameters. The automated protocol is comprised of several independent, automated “modules” that are separated by “stop points”. Modules have been designed to include a stop point at which the method can be terminated and samples stored indefinitely without affecting their chemical or activity state.



**Fig. 1** Flowchart of manual and automated methods for protein solubility screening using split-GFP. The manual approach is comprised of many individual steps that are performed separately.

### Development of the methods for the Biomek FX and SAMI method editor

All of the methods for the solubility assay have been developed for the Biomek FX workstation containing 96-multichannel head and Span-8 system. The integrated automation system contains five elements: (1) Cytomat 24 Hotel (labware storage) directly connected and controlled by Biomek FX workstation; (2) an ORCA<sup>®</sup> arm with an integrated rail facilitates transport of labware between Biomek FX workstation and different peripheral devices and instruments; (3) a DTX plate reader—for absorbance and fluorescence measurements with exchangeable filters; (4) a Cytomat 2C with Plate Shuttle System used for automated incubation of plates at different temperatures, shaking speeds and shaking directions; and (5) a Rotanta 46 RSC refrigerated robotic centrifuge used for bacterial cells centrifugation, separation of soluble and insoluble fractions, talon binding and washes. The Biomek FX workstation is

For the automated assay the procedure is divided into modules that combine several steps into one continuous process flow

controlled by Biomek software that allows building of the methods with precise control of each individual step of the method. Once the method is created it can be used on its own (Biomek FX workstation only) or incorporated into a SAMI method editor that becomes a node or module of the integrated robotics system. A SAMI method editor enables communication between different instruments comprising automated system and scheduling the automated run of the designed method. All robotics components can also be independently controlled within the SAMI method editor.

Overnight culture growth and inoculation of the starter culture

In the manual protocol colonies selected from an *in vivo* screen are grown in 3 ml of selective LB medium in 10 ml glass culture tube or 1 ml in 96-deep well plates using multichannel pipettor. Cultures are then typically grown at 30°C, shaken at 340 rpm until saturated (overnight) in the

Innova 4230 floor incubator shaker. Expression pathway starts with the inoculation of the 3.5 ml of selective LB media with 35  $\mu$ l overnight starter culture in a 10 ml glass tube, or 1 ml LB with 10  $\mu$ l of overnight starter culture (96-deep well plate) and outgrowth at 37°C, 340 rpm, until OD600 reaches 0.5 (~1.5 h). For the first automated module, colonies are grown overnight in a 96-well flat bottom tissue culture plates containing 175  $\mu$ l selective LB media supplemented with 7.5% glycerol, at 30°C, standing (not shaking), for 16 h, also in the Innova 4230 floor incubator shaker. The first automated module starts with inoculation, placing the overnight bacterial culture plates on the Biomek FX deck and transfer of 10  $\mu$ l of the overnight starter culture to 160  $\mu$ l of selective LB media in a 96-well, round bottom, tissue culture plates. The inoculated plates are then transported by the ORCA arm to the Cytomat 2C incubator and are shaken at 500 rpm for 2 h, reaching OD600 of ~0.5. All absorbance readings are performed by DTX plate reader. Critical points at this stage of the assay include a selection of the 96-well tissue culture plates for bacterial growth and choosing the optimal shaking speed in the Cytomat 2C incubator, as well as shaking direction. Considering various types of 96-well tissue culture plates available (flat, conical or round bottom), we obtained healthy bacterial growth by using round bottom plates. We have also noted the difference between shaking direction of the plates and their shaking speed. Out of 6 possible orientations (N-W-S-E, N-E-S-W, NW-SE, NE-SW, E-W and N-S) [N-North, W-West, S-South, E-East] only diagonal (NW-SE, NE-SW) direction and a rotation speed of 500 rpm resulted in healthy growth that was comparable to the manual incubation using floor incubator shaker. Other combinations of shaking direction and speed resulted in some degree of cell death. Using a “checkerboard” pattern of fluorescent sodium fluorescein solutions, and distilled water only, in alternating wells on the 96-well plate we demonstrated no cross-contamination under the optimal shaking conditions (*above*).

#### Induction, expression and quenching of the bacterial cells growth

In the manual method, the expression 3.5 ml cultures are induced with 3.5  $\mu$ l of Anhydrotetracycline 0.3 mg/ml (0.3  $\mu$ g/ml final concentration) in the 10 ml glass tube. In 96-well deep well plate the expression was induced with 100  $\mu$ l of 3  $\mu$ g/ml Anhydrotetracycline solution previously diluted in LB medium. In both cases proteins were expressed by shaking at 340 rpm at 37°C for 2 h, or 3 h at 27°C. Cells were then collected, centrifuged and resuspended in 150  $\mu$ l of TNG buffer. In the automated method, 170  $\mu$ l of inoculated cells were induced with 10  $\mu$ l of 0.6  $\mu$ g/ml Anhydrotetracycline solution diluted in

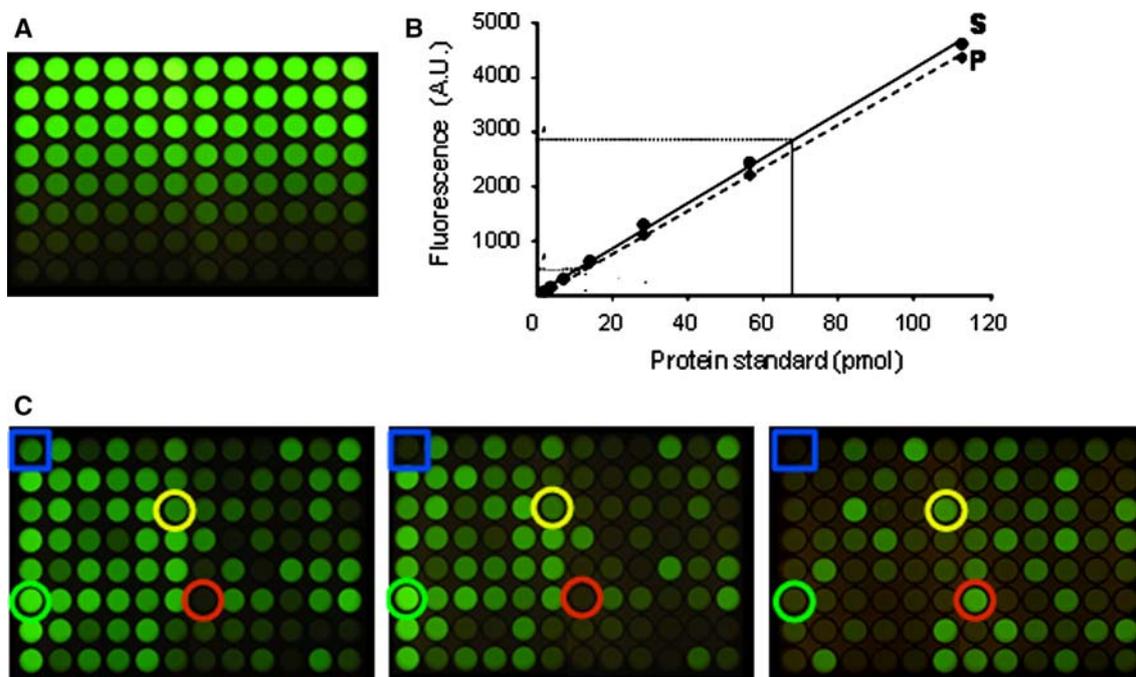
LB medium. The induction of expression was performed on the Biomek FX deck, followed by 2.5 h expression in the Cytomat 2C incubator, at 32°C, 500 rpm, NE-SW direction. Tissue culture plates with expressed proteins were then brought back to the Biomek FX deck and expression was quenched by the addition of 10  $\mu$ l of 0.5 mg/ml chloramphenicol. The quenched bacterial cultures were centrifuged, supernatant removed and the cells are frozen at -80°C. This was also the first “stop point” and completion of the first automated module (Fig. 2, Module 1).

#### Cell lysis, separation of soluble lysate from cell pellets

Cells were lysed either by sonication or chemical extraction reagent. In the manual method, disruption of the bacterial cells by chemical lysis was performed by following the manufacturer’s protocol. Individual cell preparations were lysed by sonication one-by-one (larger volumes) using a Sonicator equipped with 1/8-inch tip or in 96-well PCR plate using Sonicator—ultrasonic processor XL20-20. Following sonication, soluble and insoluble fractions are separated by centrifugation. In the automated method the previously frozen cells were either resuspended in 120  $\mu$ l of TN buffer for sonication or chemically lysed with 120  $\mu$ l of lysis reagent on the Biomek FX by addition of the chosen chemical extraction reagent or lysozyme. For sonication, TN buffer was added to 96-well PCR plates (containing frozen cells) using Biomek FX. Plates are then manually sonicated on ultrasonic processor XL20-20. Sonicated plates were returned to the Biomek FX deck, then transported by an ORCA arm to Rotanta centrifuge and centrifuged for 20 min at 4,000 rpm. Following the centrifugation the plates were again returned to Biomek FX deck and 120  $\mu$ l of supernatant (soluble fraction) was aspirated from the wells without disturbing the pellet and transferred to a fresh plate.

#### GFP 1–10 complementation of soluble and insoluble fractions

Following the separation of soluble and insoluble (pellet) fractions, 40  $\mu$ l of soluble fraction was complemented with 190  $\mu$ l of GFP 1–10 detector fragment. The remaining 80  $\mu$ l of soluble fraction were saved as a “backup” source in case of unexpected failure of any of the subsequent steps or for use in other downstream applications. The insoluble pellets can either be frozen for future processing or be washed with either TNG or TN buffer and denatured with 60  $\mu$ l of 9 M urea. Although both manual and automated methods use the same procedure, the critical points include the prevention of forming urea crystals and making sure that there is no liquid remaining after washing the pellet,



**Fig. 2** **a** Fluorescence image of 12 separate preparations of exogenous GFP 1–10 detector fragment complemented with 8 different concentrations of the control protein, sulfite reductase. **b** Calibration with control protein enables quantification of soluble protein (S) and denatured pellet (P) from fluorescence assay. **c** Fluorescence images of 96-well assay plates for final soluble fraction (left), imidazole eluted fraction (middle) and pellet fraction (right). Examples of four possible outcomes: (1) totally soluble protein that is eluted from the Talon<sup>®</sup> resin (full-length), no pellet fraction fluorescence signal is

visible for this protein (*green circle*); (2) protein that is partially soluble, the fluorescence signal is present on both final and eluted plates but a substantial proportion of the protein is insoluble (*yellow circles*); (3) totally insoluble protein, the only fluorescence signal visible is on pellet fraction plate (*red circles*); (4) protein that is soluble, but failed to bind to Talon<sup>®</sup> resin and no signal is present in pellet fraction (*blue square*), likely the protein is truncated or the 6-HIS tag is inaccessible

before the addition of urea. “Wet” pellets can result in the reduction of urea concentration that is insufficient for quantitative dissolution of the pellets. In the pellet assay only 10  $\mu$ l of the denatured sample is used for complementation with 190  $\mu$ l of GFP 1–10 and fast processing is necessary to prevent evaporation. The remaining 50  $\mu$ l of denatured pellet is saved and can be used for other applications, e.g. a refolding screen. It is also critical that following the complementation the initial fluorescence is measured as soon as possible to detect any possible GFP 1–10 leakage from the GFP 1–10 plasmid that is present in the expression strain. This is easily achieved in the automated method, where the current method allows for four 96-well plates to be assayed at any one time. Measurement of the initial fluorescence of the soluble and pellet fractions using DTX plate reader is a “stop point” between the second and third automated module (Fig. 2, Module 2).

#### Measurement of the final fluorescence of soluble and insoluble fractions

The initial rate and final value of the GFP 1–10 complementation reaction is linear over at least 4 orders of

magnitude, and samples with  $>2$  pmol protein in a 200  $\mu$ l assay reaction can be quantified within 15 min as long as calibration standards and test samples are measured at the same time [10]. Since many plates are handled, we wait at least 24 h to measure the final complementation value so that the calibration standards need be measured only once and all samples have reached the plateau values. This allows us to process hundreds of samples at any one time without affecting the final fluorescence values between the samples. After initial complementation at room temperature (ca. 20°C), plates are stored at 4°C where the GFP 1–10 is stable for several days. Once complemented, the GFP fluorescence is stable for weeks. Measurement of the final fluorescence values for soluble and insoluble fractions is a stop point between module three and four (Fig. 2, Module 3).

#### Talon binding of 6-His tagged proteins and imidazole elution

Since the S11 tag is located at the C-terminal end of the protein of interest, it is possible that proteins are expressed in truncated forms still expressing S11 tag, nonetheless

complementing with GFP 1–10 forming fluorescent GFP. The fifth automated module is used to verify full-length constructs. 100  $\mu$ l 50% Talon<sup>®</sup> resin slurry is added to complemented soluble fractions, followed by incubation, centrifugation and removal of unbound protein. The N-terminus of proteins has a 6-HIS tag, therefore only full-length constructs will remain on the Talon<sup>®</sup> resin after stringent washes. The Talon<sup>®</sup> resin-bound proteins are washed three times with TN buffer and fluorescence is read again. The bound proteins are then eluted with 300 mM imidazole (final concentration), transferred to a new NUNC Maxisorp white assay plate and the remaining fluorescence of the eluted fraction is read. The fluorescence reading in this last step corresponds to protein containing a 6-His tag. In this module the critical point is to ensure that Talon<sup>®</sup> resin is resuspended at all times to prevent blocking of the tips and thus failing the transfer of the talon resin to complemented protein fraction. The measurement of the imidazole eluted sample is the stop point of the final, fourth automated module (Fig. 2, Module 4).

#### Digital fluorescence imaging of complemented soluble and insoluble fractions

The fluorescence values obtained at the initial and final stages of complementation are used to calculate changes in fluorescence that in turn is used to calculate the number of moles of tagged protein present at these steps [10]. Digital fluorescence images allow visualization of the complementation of the assayed soluble and insoluble fractions *in vitro*. Figure 2a shows the fluorescence image of 12 different preparations of 200  $\mu$ l assay solutions containing the GFP 1–10 detector fragment complemented with 8 different concentrations of the soluble GFP S11 tagged control protein, sulfite reductase. Figure 2b shows a standard calibration that enables the quantification of the soluble protein. Similar calibrations can be performed in the presence of urea to quantify urea denatured pellets. Soluble and insoluble fractions of the assayed protein samples can be evaluated visually using fluorescence images (Fig. 2c).

#### Data collection

The automated method is designed to be able to process hundreds of samples in a single experiment. The goal of the automated solubility screening is to obtain solubility information for proteins expressed on a small scale and to be able to apply the results for making decisions regarding scale-up and large scale protein production. For quality control assurance the experimental procedure contains nine different points at which fluorescence and/or absorbance data are collected. Table 1 shows data collection events,

the information furnished. Referring to Table 1, the absorbance of the overnight cultures provides information about the state of the growth and health of bacteria that contain the plasmid coding the protein of interest. The absorbance data collected following the outgrowth and induction provides important information on the inoculation step, success or failure of induction and expression, possible toxicity effects of the target protein on bacteria. The data can be used to distinguish the clones that are healthy, grew well from the clones that did not grow at all or exhibited compromised growth. The absorbance can also be used to estimate biomass for normalization of fluorescence (see below). By measuring the initial fluorescence of the soluble and insoluble fractions prior to addition of the GFP 1–10 detection reagent, problems stemming from leakage expression of GFP 1–10 leakage from the pET plasmid present in the expression strain can be spotted and used to detect unanticipated problems with induction and expression timing. Final fluorescence data represents final complementation of the soluble and insoluble fractions, and is used to compute the fraction of the total expressed protein that is partitioned into soluble protein [10]. Talon<sup>®</sup> resin-bound and imidazole eluted fractions fluorescence distinguishes between full-length and truncated protein constructs, and is used to confirm that bound proteins can be successfully eluted. By compiling both absorbance and fluorescence data it is possible to calculate the amount of soluble protein produced per biomass of bacterial cells and therefore to identify protein constructs and conditions for scale-up of protein production, depending on the downstream applications.

#### Considerations and practical aspects for bench-to-automation

Moving the split-GFP *in vitro* assay platform from the laboratory bench to high-throughput automated platform required extensive process optimization and, in many instances, manual operations had to be re-engineered or extensively modified to make them compatible with automation.

#### Culture growth

Following overnight growth of selected colonies, the outgrowth performed on the Biomek FX and Cytomat 2C incubator had to be optimized to find the volumes and shaking protocols that simultaneously eliminated cross-contamination, and maximized volume and viable biomass production. Several tissue culture plates were tested with or without lids to optimize the growth of bacterial cultures, at the same time ensuring that there is no variability from well

**Table 1** Data collection points and information obtained at different stages of the experimental procedure

Data collection points	Information obtained
1. A <sub>600</sub> of the overnight culture	Growth and integrity of the bacterial culture containing plasmid of interest
2. A <sub>600</sub> after 2 h outgrowth	State of the bacterial cultures after inoculation, possible cross-contamination
3. A <sub>600</sub> after 2.5 h induction	Biomass production, possible toxicity, possible cross-contamination
4. Initial fluorescence of the soluble fraction	Monitoring leakage expression of GFP 1–10 detector fragment
5. Initial fluorescence of the insoluble fraction	Monitoring leakage expression of GFP 1–10 detector fragment
6. Final fluorescence of the soluble fraction before Talon <sup>®</sup> resin binding	Amount of complemented soluble protein before binding to Talon <sup>®</sup> resin
7. Talon <sup>®</sup> resin-bound fluorescence of the soluble fraction	Binding of protein to Talon <sup>®</sup> resin
8. Imidazole eluted fluorescence	Protein that successfully is eluted from Talon <sup>®</sup> resin
9. Final fluorescence of the insoluble fraction	Amount of insoluble protein and proportion of the total protein

to well in terms of cell density, change in volumes due to evaporation or shaking conditions. Although the temperatures of incubation are the same in manual and automated method, shaking speed showed to be a very important factor in automated method. Cells grown at the same speed as in the manual method (340 rpm) showed reduced growth and noticeable cell death. Maximal biomass production, comparable to the manual bench-marking method, was obtained only when the speed was increased to 500 rpm and the specific shaking direction (NW–SE) was used.

#### Induction and harvesting

Since the automated method used much smaller volumes (ca. 200 µl growth) compared to manual method (1 ml growth), the induction, expression and quenching had to be also modified. Addition of Anhydrotetracycline for induction and subsequent chloramphenicol addition for quenching of the expression resulted in increased total volume in the microplate well. It is important that all volumes are carefully calculated when planning the experiment as incorrect volume at any one step can result in failure of the whole experiment. Centrifugation of the bacterial cells after expression and discarding of the supernatant requires precise adjustment of tips movement to maximize the aspiration of the media without disturbing the pellet. It is an easier procedure in manual method, in which centrifuged samples can be turned upside down to completely remove the media and dry the pellet. At this stage we have introduced a ‘stop point’ at which the researcher has a choice of either freezing the pellet for future processing or directly proceeding to cell lysis.

#### Cell lysis and separation of soluble and pellet fractions

There are many lysis methods available and amazingly there is no single method that is commonly used by

different structural genomics consortia [13]. Our automated method allows us to choose between chemical lysis, use of lysozyme or processing of manually sonicated plates. One of critical points here is the separation of the soluble lysate from the pellet. As with the removal of the media after the expression, aspiration speed, tips movement within the wells had to be optimized to prevent carry over of the pellet with the lysate but at the same time maximizing aspiration of the lysate.

#### Complementation assay and denaturation of pellet fraction

Accurate measurement of the pellet fraction requires quantitative solubilization of the insoluble pellet fractions with urea, so inclusion body pellets needed to be dry (<5 µl buffer remaining on pellet) before dissolution in 60 µl of 9 M urea. Many steps of the split-GFP assay could in principle be scaled to arbitrarily small volumes, but during the processing of 4 or more 96-well plates for urea-dissolved pellet fractions at the concurrently, any volume of urea less than 10 µl resulted in urea crystal formation due to partial evaporation of the sample. Complementation with GFP 1–10 detector fragment is time-dependent process, and is very rapid for concentrated samples. It is important to measure the initial fluorescence (i.e., the background fluorescence) immediately after adding the GFP 1–10 reagent complementation to avoid including authentic complementation signal in the background fluorescence value that will be subtracted from the final complementation reading. Following final fluorescence reading of soluble fraction, Talon<sup>®</sup> resin is added to complemented soluble fraction to assay for the full-length, 6-HIS tag containing constructs. To implement this step on the Biomek FX deck, the 50% Talon<sup>®</sup> resin slurry must be uniformly resuspended to prevent blocking of the tips and uniform bead/resin transfer. Talon<sup>®</sup> resin-bound proteins

are eluted with imidazole and transferred to a new assay plate for fluorescence reading.

## Discussion and conclusions

The split-GFP system is stable and sensitive, and is a flexible tool with numerous applications. For structural genomics, predicting protein solubility using self-assembling, complementary fragments of the GFP is an alternative method to many conventional approaches based on SDS-PAGE, antibody blots or commercially available “split” systems. Split-GFP can be used both in vivo and in vitro, does not require exogenous reagents, is very sensitive and is, as shown here, amenable to high-throughput robotics automation. Protein solubility screening using split-GFP includes several steps that previously have been optimized for manual sample manipulation. Here we presented the transferring of the manual split-GFP solubility assay into an automated, robot-based system using Biomek FX liquid-handling robot coupled with Cytomat incubators, DTX plate reader, Rotanta centrifuge and ORCA transportation rail. These are the minimal requirements to implement our protocol. There are many liquid-handling platforms, different incubators, centrifuges, and plate readers on the market that constitute integrated robotics systems that can adopt our method. All our SAMI automated modules, as well as individual Biomek FX methods can be found at our website—<http://www.lanl.gov/projects/gfp/Solubility.shtml>. For all Biomek FX and SAMI users these can be adopted with some modifications for specific Biomek FX platforms and for users of other robotics systems should become a reference point for building corresponding methods. During the automation procedure we have identified several critical steps that had to be optimized for the robotic platform. These include type of labware used for incubation, shaking speed and orientation, minimum volumes that can be handled without evaporation or changes in volumes between the wells. We have eliminated laborious tasks of manual transfer of samples between assay and analytical plates, delivery of small volumes of expression induction and quenching reagents, as well as delivery of Talon<sup>®</sup> resin for 6-HIS tag containing soluble protein fractions. Implementation of the robotic workstation is especially useful when hundreds of protein constructs have to be handled in parallel with reproducible results. In summary, automated high-throughput approaches for multiple protein sample handling are becoming more and more important, as most of the “low-hanging” protein targets have been already selected and characterized. As we reach higher and higher for the new targets, we face new challenges. Proteins are complex, difficult to work with in their native form, require

binding partners or other factors to be stable or functional. Finding these optimal constructs and conditions is a time-consuming challenge requiring high-throughput methodologies and assays. Split-GFP is not only a useful tool for a rapid solubility screening of the full-length cDNAs but it can also be applied to assess and delineate the solubility of individual protein domains by library screening methods. This is especially useful when applied to biologically important proteins that initially intractable as a research reagent or require a traditional laboratory approach that often takes years to complete with uncertain outcome. The establishment of high-throughput methodologies is especially useful when identification of stable form of protein or individual domain requires screening of thousands of clones from a single library. The split-GFP strategy coupled with automated, high-throughput platform presents a powerful method that can be used in salvaging thousands of “difficult to handle” proteins, many of which could be of medical importance or serve as new therapeutic targets. Currently our robotics resources allow us to physically process 24 96-well plates per week. Data collected at different stages of the assay is compiled into one place and analyzed manually, a rate limiting step. Our near future goal is to implement an automatic data cataloging/barcoding and analysis system for thousands of protein samples. Our goal is to further increase the throughput by moving into 384 well plates and incorporate a “refolding” module for well-expressed but insoluble targets.

**Acknowledgments** The authors wish to acknowledge Jean-Denis Pédelacq and Carolyn Bell for helpful comments. This work was supported by the NIH Protein Structure Initiative (U54 GM074946-01US, “Integrated Center for Structure and Function Innovation”).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Knaust RK, Nordlung P (2001) Screening for soluble expression of recombinant proteins in 96-well format. *Anal Biochem* 297:79–85. doi:[10.1006/abio.2001.5331](https://doi.org/10.1006/abio.2001.5331)
2. Vincentelli R, Canaan S, Offant J, Cambillau C, Bignon C (2005) Automated expression and solubility screening of His-tagged proteins in 96-well format. *Anal Biochem* 346:77–84. doi:[10.1016/j.ab.2005.07.039](https://doi.org/10.1016/j.ab.2005.07.039)
3. Nguyen H, Martinez B, Oganessian N, Kim R (2004) An automated small-scale protein expression and purification screening provides beneficial information for protein production. *J Struct Funct Genomics* 5:23–27. doi:[10.1023/B:JSFG.0000029195.73810.86](https://doi.org/10.1023/B:JSFG.0000029195.73810.86)
4. Wigley WC, Stidham RD, Smith NM, Hunt JF, Thomas PJ (2001) Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat Biotechnol* 19:131–136. doi:[10.1038/84389](https://doi.org/10.1038/84389)

5. Horecka J, Charter NW, Bosano BL, Fung P, Kobel P, Peng K, Eglén RM (2006) Antibody-free method for protein detection on blots using enzyme fragment complementation. *Biotechniques* 40:381–383. doi:[10.2144/000112119](https://doi.org/10.2144/000112119)
6. Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, Llopis J, Tsien RY (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J Am Chem Soc* 124:6063–6076. doi:[10.1021/ja017687n](https://doi.org/10.1021/ja017687n)
7. Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein folding assay using green fluorescent protein. *Nat Biotechnol* 17:691–695. doi:[10.1038/10904](https://doi.org/10.1038/10904)
8. Bertens P, Heijne W, Van Der Wel N, Wellink J, Van Kammen A (2003) Studies on the C-terminus of the Cowpea mosaic virus movement protein. *Arch Virol* 148:265–279. doi:[10.1007/s00705-002-0918-z](https://doi.org/10.1007/s00705-002-0918-z)
9. Cabantous S, Terwilliger TC, Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 23:102–107. doi:[10.1038/nbt1044](https://doi.org/10.1038/nbt1044)
10. Cabantous S, Waldo GS (2006) In vivo and in vitro protein solubility assays using split-GFP. *Nat Methods* 3:845–854. doi:[10.1038/nmeth932](https://doi.org/10.1038/nmeth932)
11. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a super folder green fluorescent protein. *Nat Biotechnol* 24:79–88. doi:[10.1038/nbt1172](https://doi.org/10.1038/nbt1172)
12. Cabantous S, Pédelacq JD, Mark B, Naranjo C, Terwilliger TC, Waldo GC (2005) Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J Struct Funct Genomics* 8:153–166
13. Berrow NS, Büssov K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentell IR, Busso D (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr D Biol Crystallogr* 62:1218–1226. doi:[10.1107/S0907444906031337](https://doi.org/10.1107/S0907444906031337)