



# Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants

Sandra K. Tanz<sup>1\*</sup>, Ian Castleden<sup>2</sup>, Ian D. Small<sup>1,2</sup> and A. Harvey Millar<sup>1,2,3</sup>

<sup>1</sup> ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, Perth, WA, Australia

<sup>2</sup> Centre of Excellence in Computational Systems Biology, The University of Western Australia, Perth, WA, Australia

<sup>3</sup> Centre for Comparative Analysis on Biomolecular Networks (CABiN), The University of Western Australia, Perth, WA, Australia

## Edited by:

Katja Baerenfaller, Swiss Federal Institute of Technology Zurich, Switzerland

## Reviewed by:

Martin Hajduch, Slovak Academy of Sciences, Slovakia  
Borjana Arsova, Heinrich-Heine University, Germany

## \*Correspondence:

Sandra K. Tanz, ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia  
e-mail: sandra.tanz@uwa.edu.au

Fluorescent protein (FP) tagging approaches are widely used to determine the subcellular location of plant proteins. Here we give a brief overview of FP approaches, highlight potential technical problems, and discuss what to consider when designing FP/protein fusion constructs and performing transformation assays. We analyze published FP tagging data sets along with data from proteomics studies collated in SUBA3, a subcellular location database for Arabidopsis proteins, and assess the reliability of these data sets by comparing them. We also outline the limitations of the FP tagging approach for defining protein location and investigate multiple localization claims by FP tagging. We conclude that the collation of localization datasets in databases like SUBA3 is helpful for revealing discrepancies in location attributions by different techniques and/or by different research groups.

**Keywords:** FP tagging, subcellular localization, database, Arabidopsis, subcellular proteomics

## INTRODUCTION

Plant systems are comprised of a complex network where organs, tissues, and cell types interact with each other. Each cell, in turn, is characterized by a comparably complex network of subcellular compartments that are morphologically and functionally different. Proteins located in these subcellular compartments often share similar attributes and play roles in defining the function of these distinct cellular environments. To understand how plant cells are functionally structured, we need to know where enzymes and regulatory proteins are located within the cell at certain points in development and under particular environmental conditions (Millar et al., 2009).

Different methods can be employed to help to determine a protein's intracellular location. Computational programs that can predict the subcellular location from the protein's nucleic acid sequence are useful but not conclusive (Richly and Leister, 2004; Heazlewood et al., 2005; Reumann, 2011). In addition, some proteins exist in multiple locations (Small et al., 1998; Carrie and Small, 2012) but only a few prediction programs deal with multiple locations effectively, such as ATP (Mitschke et al., 2009), Plant-mPLoc (Chou and Shen, 2010), WOLF PSORT (Horton et al., 2007), and YLoc (Briesemeister et al., 2010; for an overview of protein localization predictors see also Tanz and Small, 2011). *In vitro* uptake studies of an exogenously added protein into an isolated organelle has been a powerful tool for detailed studies of the import process but does not reproduce the complex intracellular environment and might not always reveal targeting preference between organelles (Rudhe et al., 2002; Chew et al., 2003). Immunolabeling of proteins in tissue sections, where specific antibodies recognize the native conformation of the protein, can be laborious and time-consuming and may not always be successful. This approach is also problematic when dealing

with proteins with closely related sequences. Proteomic studies employing cell fractionation and mass spectrometry (MS) to identify peptides in the purified subcellular compartments result in large, information-rich datasets (Jaquinod et al., 2007; Reumann et al., 2007, 2009; Eubel et al., 2008; Mitra et al., 2009; Ferro et al., 2010; Olinares et al., 2010; Ito et al., 2011; Klodmann et al., 2011; Lee et al., 2011; Taylor et al., 2011; Zhang and Peck, 2011; Lundquist et al., 2012). However, MS can be technically challenging as contamination of the subcellular preparation with proteins from other parts of the cell is a frequent problem and low abundance, small and hydrophobic proteins can be missed employing this approach. Fusion of fluorescent protein (FP) coding sequences to the coding regions of genes of unknown location is relatively simple and fast and can be directed to specific proteins of interest, and as a result FP tagging has become the method of choice for many plant biologists.

FP tagging and subcellular proteomic studies have become the dominant tools for determining the location of a protein within the plant cell and provide complementary and independent information. However, these high-throughput approaches are prone to both false-negative and false-positive claims of protein location. In addition, the FP tagging approach defines a protein's targeting ability and defines a final location by accumulated fluorescent signal, while the subcellular proteomics approach determines, in steady-state, where the native protein accumulates in the cell. While it is expected that these two approaches should reveal matching results in most cases, they will not always agree even when the data from both methods is sound (Millar et al., 2009). Collating location data sets of different approaches in databases like SUBA (Heazlewood et al., 2007; Tanz et al., 2013) allows users to assess these data collectively and can expose discrepancies and conflicts in location attributions by different methods and/or by

different research groups. In this report we review the current location data sets in SUBA3 (Tanz et al., 2013). Specifically, we focus on the subcellular location data by FP tagging and examine the broader reliability of these data compared to other experimental claims, discuss the limitations of the approach, and analyze localization claims by FP for the same protein in multiple locations.

## THE FP TAGGING APPROACH

### FP TAGGING IN PLANTS

Expression of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* and its spectral variants within cells (Chalfie et al., 1994; Zacharias and Tsien, 2006) has stimulated many experiments to gain new insights into the organization of cellular metabolism and to better understand compartmentation of cells. FP tagging can now provide answers to the following questions: Where do proteins localize within the cell? Where do dynamic proteins move within the cell? How do individual proteins behave in response to developmental and environmental changes? However, heterologous expression of GFP in plant cells has not always been straightforward. Initially, GFP tagging was only successful in animal and fungal cells, whereas only poor GFP expression levels were observed in plant cells. This was due to the presence of a cryptic intron in the original jellyfish GFP sequence, which was incorrectly removed in plant systems. Modifications to the GFP codon sequence abolished the erroneous removal of part of the sequence and restored the expression of GFP in plant systems (Haseloff et al., 1997; Rouwendal et al., 1997).

Today, GFP and its derivatives and homologs (here collectively referred to as fluorescent proteins or FPs) are the most important fluorophores for plant cell biology and their use has been reported extensively in the literature (reviews include Hanson and Kohler, 2001; Ehrhardt, 2003; Dixit et al., 2006; Fricker et al., 2006; Berg and Beachy, 2008). Untargeted or “free” FPs are localized to the cytoplasm in plant cells but also go into the nucleus due to their small size. In addition, FPs have been targeted to all plant organelles using FP fusions incorporating location-specific signal sequences (Tian et al., 2004). In fact, a set of fluorescent organelle markers has been generated based on well-established targeting sequences (Nelson et al., 2007). All markers were generated with four different FPs in two different binary plasmids to allow for flexible combinations during co-localization studies (Nelson et al., 2007). The use of FPs to localize individual proteins is based on the ability to engineer FP fusions, with FP tagged onto the protein of interest, allowing it to be observed within intact tissue. FPs have even been used to tag viral proteins to investigate the interaction of such proteins with plant organelles (Lazarowitz and Beachy, 1999; Ueki and Citovsky, 2011). FP imaging does not require staining and allows analysis of cells in a relatively undisturbed, living state. This non-invasive way of monitoring localization and dynamics of proteins as well as there being no need for exogenous substrates or co-factors (Chalfie et al., 1994) are the main advantages of FP tagging.

A disadvantage with FP imaging, particularly in plants, has been the autofluorescence of cellular components such as cell walls and plastids, which may overlap with FP spectral signals (Deblasio et al., 2010). For example, interference by

autofluorescence from the cell wall could be a problem for the localization of low abundant plasma membrane proteins. However, most modern confocal microscopes are now able to account for background autofluorescence and subtract it from FP signals based on the unique spectral profile of non-FP expressing reference images.

As increasing numbers of plant genomes are fully sequenced, high-throughput FP screens are being employed to identify gene function and regulatory networks (Cutler et al., 2000; Escobar et al., 2003; Tian et al., 2004; Koroleva et al., 2005; Marion et al., 2008). For example, a library of Arabidopsis cDNAs was generated and fused to the 3' end of GFP. The library was then transformed into Arabidopsis *en masse* and the progeny screened for transgenic plants showing different subcellular localization patterns (Cutler et al., 2000). In a complementary study, open reading frame cDNA clones were GFP-tagged at their 3' end and transformed cell cultures were screened for localization patterns (Koroleva et al., 2005). The Arabidopsis localizome project uses a recombineering-based gene tagging approach to generate FP fusion proteins in their chromosomal context (Zhou et al., 2011). A bacterial homologous recombination system is used to insert FP tags into genes of interest that are harbored by transformation-competent bacterial artificial chromosomes (TAC; Zhou et al., 2011). This ensures that all *cis*-regulatory sequences of a gene are included and because the genes are not amplified by PCR there is no limit to the size of a gene that can be tagged. Thus, this is a promising approach for the future that will eliminate many of the current problems encountered during FP tagging studies (see section Considerations with FP/Protein Fusions).

### CONSIDERATIONS WITH FP/PROTEIN FUSIONS

The fusion of FP to enzymes often does not inhibit their catalytic activity and FP tagging is generally thought to be a “safe method” to determine the subcellular location of a protein. Indeed expressions of FP fusions of proteins have been reported to functionally complement knockout mutants (Sedbrook et al., 2002; Benkova et al., 2003; Kim et al., 2003). However, it is possible that in some cases the FP/protein fusion and the wild-type protein will differ in their subcellular locations leading to false positive results. Careful consideration is required where a protein is tagged, as the presence of the FP could hinder proper localization encoded by a transit sequence on the attached protein.

FP coding sequences are typically fused to either the 5' or 3' end of the coding region of a DNA sequence in question, generating N- or C-terminal FP fusions (Cutler et al., 2000; Huh et al., 2003). Alternatively, proteins can be tagged at a selected internal site, which has the advantage that targeting signals present at the 5' or 3' end of the coding region are not masked by the FP. For example, N-terminal fusions (FP is fused to the N terminus of the protein of interest) interfere with plastid and mitochondrial localization signals and are also likely to abrogate endoplasmic reticulum (ER) signal peptides. C-terminal fusions (FP is fused to the C terminus of the protein of interest) may also cause many proteins to mislocalize, particularly peroxisomal proteins. In addition, C-terminal fusions could mask stem-loop structures in the 3' part of the coding sequence and the 3' untranslated region, which are necessary for the accurate localization

of certain mRNAs (Chartrand et al., 1999). N- or C-terminal fusions may also interfere with posttranslational modification sites, such as myristylation or farnesylation sites important for membrane targeting. Indeed, some plasma membrane proteins failed to localize to the plasma membrane using N- or C-terminal tags but internally tagged proteins localized correctly (Sedbrook et al., 2002; Gardiner et al., 2003; Tian et al., 2004). In addition, more and more multi-targeted proteins are being identified. For example, proteins with peroxisomal targeting signals and chloroplast or mitochondrial transit peptides have only been identified when analyzed with separate N- and C-terminal fusion constructs (Carrie et al., 2008; Hooks et al., 2012). Thus, for correct localization it is crucial to examine N- and C-terminal FP fusion constructs and/or internally tagged proteins.

Similarly, the length of a protein sequence for fusion with an FP needs to be considered. Using the full-length sequence of a protein is desirable; however, some genes might be too long to be easily cloned into an expression vector and thus partial sequences are frequently used for localization by FP tagging. Most plastid or mitochondrial targeting sequences are located at the N-terminus and the N-terminal ~100 amino acids are generally sufficient for correct subcellular localization. However, in this case a possible second C-terminal or internally located targeting sequence might be missed, as in the case of multi-targeted proteins (Carrie et al., 2009; Hooks et al., 2012).

The promoter used in front of an FP fusion construct also needs to be considered. Often the CMV 35S promoter is used instead of the native gene promoter, which could lead to higher expression levels of the fusion construct than for the endogenous protein, and subsequently could lead to mistargeting. This could particularly affect nuclear-encoded proteins targeted to organelles, where high protein abundance could result in incomplete import. Theoretically this might also account for some false claims of dual targeting of proteins between the cytoplasm and various organelles.

In addition, the fused FP could be the reason for a conformational modification in the attached protein and a localization signal could become active, which is normally isolated in the absence of FP or when it is lacking some endogenous ligand. Also, the abundance of the fused FP may be very different from the native protein, leading to mislocation, aggregation, metabolic disturbance or the like.

### CONSIDERATIONS WITH TRANSFORMATION ASSAYS DURING FP TAGGING

FP fusion constructs can be introduced into plant cells for transient assays or stably expressed in transgenic plants. With the latter, many different cell types can be investigated in which the FP/protein fusion is expressed, while not all cell types are suitable for transient expression. In addition, cell damage often occurs during DNA uptake in transient assays and inconsistent amounts of FP fusion constructs can be delivered into the cells. Thus, it is more reliable overall to analyse healthy stable transformants to define protein location by FP. However, the simplicity and speed of transient assays makes them a very valuable tool, especially when considering the extra labor and analysis it takes to generate and test stable transgenic plants. Onion epidermis is a

favorite material for biolistic transient assays, because of its clear cytoplasm and single layer of living cells. Similarly, *Arabidopsis* cell culture, *Arabidopsis* seedlings and young detached leaves have also been successfully used in transient assays. Following particle bombardment with various constructs, cellular compartments such as ER, Golgi, vacuole, mitochondria, plastids and plasma membrane can all be labeled by different transiently expressed FP fusions in *Arabidopsis* (Nelson et al., 2007). Other popular transient expression methods include the protein expression in isolated protoplasts by electroporation or using polyethylene glycol (Miao and Jiang, 2007; Yoo et al., 2007) and the *Agrobacterium*-mediated infiltration in *Nicotiana benthamiana* (Yang et al., 2000) or *Arabidopsis* leaves (Tsuda et al., 2012).

## ANALYSIS OF FP TAGGING DATA IN SUBA3

### THE RELIABILITY OF FP LOCALIZATION DATA

Given that various approaches have been used to define the location of proteins, and each has its own drawbacks, it is important to ask: What is the reliability of the FP tagging approach? In an attempt to answer this question we have analyzed subcellular localization data in SUBA (Heazlewood et al., 2007; Tanz et al., 2013). At the time of writing, SUBA3 contains a total of 3788 entries based on FP tagging studies from 1074 different publications, representing 2477 unique proteins. Of these, 443 proteins have been localized at least twice independently by FP, and for 375 proteins the independent FP localizations agree. Thus, for 85% of cases, the FP data are internally consistent, whereas they disagree in the cases of 123 proteins (28%). For 13% of proteins, the FP localization of one publication has been shown to agree with a second publication, and shown to disagree with a third publication; these proteins count toward both groups. Additional data based on subcellular MS-based proteomics from 122 different publications add 22,191 entries on 7685 distinct proteins. Calculating the percentage of FP tagging and MS agreements/disagreements for proteins for which both FP tagging and proteomics data are available shows that 61% of the data agree and 39% disagree. The remaining 1593 FP entries are not confirmed nor do they disagree with MS data because no independent subcellular proteomics data relating to these proteins have been published to our knowledge. Analyzing the FP data set further and comparing it to data from subcellular MS-based proteomics reveals that 849 out of 2996 FP protein claims agree with proteomics data (Table 1). The number of protein claims (2996) is different to the number of unique proteins (2477) because it includes cases where the same protein has been found in multiple compartments and thus accounts for multiple entries, and it is also different to the total FP entries (3788) as a protein is only counted once per location regardless how many researchers have found it in the same location. In these 849 cases, the protein's targeting ability tested by FP tagging agrees with the protein's accumulation tested by subcellular MS and we can be confident of the location claim and how the protein got there. On the contrary, for 554 FP claims a different location has been reported by MS studies. Thus, published disagreement of subcellular location exist for these FP claims and the protein's targeting ability appears to disagree with the claimed location of the protein's accumulation.

**Table 1 | Number of localizations by FP tagging for each of the 11 subcellular compartments in SUBA3.**

Intracellular location	FP localizations	FP localizations confirmed by MS	FP localizations contradicted by MS	FP and MS localizations agree/disagree (% of proteins with both FP and MS data)	FP localizations neither confirmed nor contradicted
Cytoskeleton	63	0	24	0/100	39
Cytosol	514	119	101	54/46	294
ER	189	31	56	36/64	102
Extracellular	40	9	10	47/53	21
Golgi	145	42	57	42/58	46
Mitochondrion	312	90	54	63/37	168
Nucleus	779	95	149	39/61	535
Peroxisome	130	59	33	64/36	38
PM	245	122	17	88/12	106
Plastid	486	248	34	88/12	203
Vacuole	93	34	18	65/34	41
Total	2996	849	554	61/39	1593

Also shown are the numbers of FP localizations that overlap with MS localization and are thus confirmed by this approach, the numbers of FP localizations that disagree with MS localizations, the percentage of agreements and disagreements of FP with MS localizations for proteins for which FP tagging and proteomics data are both available, and the numbers of FP localizations that are neither confirmed nor contradicted by MS data. Data sets were extracted from the SUBA3 database (<http://suba.plantenergy.uwa.edu.au>).

Abbreviations: FP, fluorescent protein; MS, mass spectrometry; ER, endoplasmic reticulum; PM, plasma membrane.

A detailed list of the existing FP data for each of the 11 compartments in SUBA3 is shown in **Table 1**, along with the independent confirmations and disagreements by published subcellular proteomics data. For most of the compartments, the agreements between the claims for localization by FP tagging and subcellular MS lie between 36% and 65% for proteins with both FP and MS data available (**Table 1**). However, for two compartments, namely plastid and plasma membrane, 88% of proteins for which FP and MS data are available show an agreement and only 12% of FP data do not agree with the MS localization data (**Table 1**). The relatively high discrepancy between FP and MS data for most of the other compartments (35–64%, **Table 1**), likely highlights technical problems in false positive rates with both the MS and FP tagging approaches but further analysis will be required to confirm this.

The three organelles plastid, mitochondrion and peroxisome were chosen as examples to closer investigate the proteins for which a disagreement between FP and MS data has been observed.

### Plastid

A total of 486 proteins have been localized to the plastid by FP tagging (**Table 1**). From these, the published plastid FP localizations of 34 proteins appear to disagree with the locations claimed by proteomics studies (Supplementary Table 1). For eight of these proteins, additional FP location data for the same proteins agree with MS location claims and thus the whole FP data set does not strictly disagree with the proteomics (Supplementary Table 1, AGIs with asterisk). Investigating the 34 proteins more closely reveals that seven proteins are known to be dual-targeted or dynamic so here the two data sets may both be correct (Supplementary Table 1, yellow). Another eight proteins clearly have a function in the plastid with two of these located in a

second compartment other than the one determined by MS (Gao et al., 2003; Lurin et al., 2004; Murcha et al., 2007; Yu et al., 2008; Sun et al., 2010; Skalitzky et al., 2011). Thus, the disagreements are due to technical issues with the MS approach and could result from contamination of these proteins in sample preparations of other subcellular structures (Supplementary Table 1, blue). One of these proteins is OEP16 (At4g16160), localized by FP tagging to the plastid and by MS to the cytosol, but it has been confirmed by *in vitro* imports to be targeted to plastids and not to mitochondria, unlike the mitochondrial isoforms of this protein family (Murcha et al., 2007). The disagreement is likely due to be an error or contamination in the MS approach (Supplementary Table 1, blue). One protein (Complex I subunit At2g02510) clearly functions in the mitochondrion (Brugiere et al., 2004; Meyer et al., 2008; Klodmann et al., 2011), and the disagreement in localization is due to technical issues with the FP tagging approach (Supplementary Table 1, green). These include artifacts that may result from the foreign passenger protein affecting the targeting ability of the protein of interest, such as difference in abundance of the fusion protein, conformational changes or activation of a localization signal in the attached protein (see section Considerations with FP/Protein Fusions). The remaining 18 proteins are either unknown multi-targeted proteins located to the plastid and other compartments in the cell or the disagreement between FP and MS data is due to limitations of one or both approaches.

An interesting example for when experimental data appear to disagree but when in fact they actually complement each other is alanyl-tRNA synthetase (At1g50200). FP tagging studies found this protein to be targeted to plastids and mitochondria, whereas proteomics studies found it in the cytosol (Supplementary Table 1). Analysis of the transcription of the gene showed the presence of two translation initiation codons

(Mireau et al., 1996). Translation from the upstream AUG generates an N-terminal extension with features that target the protein to the mitochondrion and plastid, whereas most ribosomes initiate on the downstream AUG to give the shorter polypeptide corresponding in size to the cytosolic enzyme (Mireau et al., 1996). Examining the peptides identified in the cytosolic MS study (Ito et al., 2011) showed that all the cytosolic peptides significantly matching to At1g50200 (see Ito et al., 2011; Supplementary Table 1, protein hit number 68) are downstream of the second start methionine. Thus, alanyl-tRNA synthetase is only expressed at low levels in mitochondria and plastids, which explains why MS studies have not found it in these organelles but only in the cytosol and why FP studies, using the full-length sequence, have only found it in plastids and mitochondria but not in the cytosol.

### **Mitochondrion**

Examining the 54 proteins that have been localized to the mitochondrion by FP tagging but elsewhere by subcellular MS studies shows that as many as 37 of these have additional FP data that agree with MS locations (Supplementary Table 1, AGIs with asterisk). Twenty six of these 54 proteins are known dual-targeted or dynamic proteins (Supplementary Table 1, yellow). In both cases no strict disagreement exists. Eight proteins are clearly localized to and have a function in the mitochondrion as defined by FP tagging (six of these are additionally targeted to a second compartment different to the one defined by MS) and the location disagreements are due to technical issues with the MS approach (Supplementary Table 1, blue) (Souciet et al., 1999; Escobar et al., 2003; Michalecka et al., 2003; Duchene et al., 2005; Murcha et al., 2007; Carrie et al., 2008, 2009; Palmieri et al., 2009). Another seven proteins are clearly not located in the mitochondrion but function in the plastid (Hjelmstad and Bell, 1990; Froehlich et al., 2003; Asano et al., 2004; Chew et al., 2004; Friso et al., 2004; Kleffmann et al., 2004; Peltier et al., 2004; Giacomelli et al., 2006; Peltier et al., 2006; Rutschow et al., 2008; Zybailov et al., 2008; Ferro et al., 2010; Olinares et al., 2010; Granlund et al., 2011), and here the disagreement in location is due to technical issues with the FP tagging approach (Supplementary Table 1, green). The remaining 13 proteins are either unknown multi-targeted proteins or the disagreement is due to limitations of the FP tagging or the subcellular MS approach.

### **Peroxisome**

One hundred and thirty proteins are localized to the peroxisome by FP tagging, of which 33 are localized elsewhere by proteomic studies (Table 1). Eight of these have additional FP data that agree with MS locations (Supplementary Table 1, AGIs with asterisk). Eight of the 33 proteins are known to be dual-targeted or dynamic proteins and the two data sets do not necessarily disagree (Supplementary Table 1, yellow). Three proteins are clearly localized to the peroxisome and have a function in the peroxisome (Cutler et al., 2000; Carrie et al., 2008, 2009) as defined by FP tagging [with two of them, a substrate carrier (At3g55640) and a NAD(P)H dehydrogenase (At4g28220), also localized to another compartment different to the one determined by MS], and the location disagreement is due to technical issues with the

MS approach (Supplementary Table 1, blue). Four proteins are either unknown multi-targeted proteins or the location difference is due to limitations of one or both approaches (Supplementary Table 1, no color). However, about half of the location discrepancies between the two methods are due to technical issues with the FP tagging approach as most proteins are most likely not localized to the peroxisome and have functions elsewhere in the cell (Supplementary Table 1, green).

### **MULTIPLE LOCALIZATION CLAIMS BY FP TAGGING**

The redundancy that is apparent between 2996 FP localizations in Table 1, but 2477 unique proteins localized by FP tagging, is either due to multiple locations claimed by single literature reports or independent reports claim different locations for a single protein. Examples for the former include dual-targeted proteins to chloroplasts and mitochondria (Peeters and Small, 2001; Carrie and Small, 2012), to mitochondria and peroxisomes (Carrie et al., 2009), and to mitochondria and nucleus (Carrie et al., 2009; Hammani et al., 2011).

Analyzing only the FP tagging data in SUBA3 generated a total of 739 claims where proteins are localized to two different locations (Table 2). The 739 claims comprise 545 distinct proteins that have been localized to at least two different cellular compartments by FP tagging. A paired matrix of these data displays these dual localization claims for each possible subcellular compartment combination (Table 2). There is typically 1–20% overlap between any two subcellular proteomes. However, a 31% and 46% overlap exists between nucleus and cytosol and a 20% and 32% overlap between plastid and mitochondrion (Table 2). This can be partially explained by dynamic proteins that can move between nucleus and cytosol and proteins that are dual-targeted to these compartments. No doubt, the FP tagging approach has its limitations and some false positive results must also be contributing to these overlaps. Furthermore, a dual localization to the nucleus and cytosol can be due to FP artifacts, including GFP localizing by itself to the cytosol and the nucleus, which can generate false positive results to these two compartments.

Of the 739 claims where proteins are localized to two different locations, 80% (595 dual claims) are by the same literature reports. These comprise 491 proteins and because the dual location is reported by the same publication these are presumably dual- or multi-targeted proteins. 20% of these claims (representing 105 proteins) demonstrate a conflict in the literature (as they appear as different publications that contradict each other) and may highlight problems associated with the use of different FP tagging approaches. However, this set could also include biological discoveries such as identification of an unknown dual-targeted protein or showing dynamic proteins that move around in the cell in different cell types or treatments.

As examples for further investigation, the dual FP localization claims for mitochondrion/plastid, mitochondrion/peroxisome, and plastid/peroxisome were chosen.

### **Mitochondrion and plastid**

Examining the literature references of the 100 proteins that have been located by FP tagging to the plastid and mitochondrion (Table 2) reveals that the dual localizations of 92 proteins are

**Table 2 | A paired matrix showing dual FP localization claims for each possible subcellular compartment combination.**

FP localization	Cytoskeleton	Cytosol	ER	Extracellular	Golgi	Mitochondrion	Nucleus	Peroxisome	PM	Plastid	Vacuole
<b>Cytoskeleton</b>	63										
<b>Cytosol</b>	11	514									
<b>ER</b>	3	15	189								
<b>Extracellular</b>	1	3	1	40							
<b>Golgi</b>	3	21	28	2	145						
<b>Mitochondrion</b>	2	12	4	0	2	312					
<b>Nucleus</b>	12	238	13	4	1	8	779				
<b>Peroxisome</b>	2	11	2	0	1	10	5	130			
<b>PM</b>	8	45	21	14	22	3	23	0	145		
<b>Plastid</b>	2	16	10	0	2	100	7	8	3	486	
<b>Vacuole</b>	1	5	10	1	10	0	2	2	7	2	93

In total, 739 claims are listed, comprising 545 distinct proteins that have been localized to at least two different cellular compartments by FP tagging. The matrix diagonal shows the set of proteins claimed in each compartment. In the matrix below the diagonal, the two-way comparisons of claims for proteins to be present in different compartments are shown. Data sets were extracted from the SUBA3 database (<http://suba.plantenergy.uwa.edu.au>).

Abbreviations: FP, fluorescent protein; ER, endoplasmic reticulum; PM, plasma membrane.

described in the same literature reports and these proteins are presumably dual-targeted (Supplementary Table 2, “Y”). Indeed when investigating the function of these proteins, many are known dual-targeted proteins (Supplementary Table 2, yellow). Nevertheless, four proteins are likely to be only located to the mitochondrion (Supplementary Table 2, orange) and another eight only located in plastids (Supplementary Table 2, green). Thus, here the apparent dual location is due to technical issues with the FP tagging approach that could involve a difference in abundance of the fusion protein or conformational changes leading to activation of a localization signal in the attached protein (see section Considerations with FP/Protein Fusions). For eight proteins a literature conflict exists and independent reports claim mitochondrial and plastid locations for a single protein. These proteins are either dual-located proteins, or the dual localizations are false positives due to technical problems with the FP tagging approach. In fact, based on their function and from independent literature reports, two of these eight proteins are already known dual-located proteins [dynamin 3A (At4g33650) and lon1 protease (At5g26860); Supplementary Table 2, yellow] and four are known to be located in the plastid only (Supplementary Table 2, green) indicating an issue with the FP approach.

#### **Mitochondrion and peroxisome**

Ten proteins have been localized to mitochondria and peroxisomes by FP tagging (Table 2) and the dual-locations of all ten proteins are each reported by the same publication, indicating all ten proteins are probably truly dual-targeted (Supplementary Table 2, “Y”). In fact, more than half of the proteins are known dual-targeted proteins from other literature (Supplementary Table 2, yellow).

#### **Peroxisome and plastid**

Of the eight distinct proteins that have been localized to the peroxisome and plastid by FP tagging, five proteins are presumably dual-targeted (same publication; Supplementary Table 2, “Y”),

of which two are known dual-targeted proteins based on the function (Supplementary Table 2, yellow). The remaining three proteins demonstrate a conflict in the literature (Supplementary Table 2, “N”), of which two are clearly only located in the plastid [Rubisco small chain 1A (At1g67090; Parry et al., 2003) and chaperonin 20 (At5g20720; Carrie et al., 2009)] and the multiple localizations of these proteins likely represent technical problems with the FP tagging method (Supplementary Table 2, green). The third is the same dynamin 3A (At4g33650) noted above; the plastid claim for this protein by FP pre-dated the dual-targeting claim in mitochondria and peroxisomes by 6 years. While an explanation of why a plastid FP location was found has not been provided, the weight of genetic and other evidence appears to suggest this is a technical problem with the FP claim of the plastid location (Mano et al., 2004).

## **CONCLUSIONS**

FP tagging with its rapidity and simplicity has become a very important tool for plant biologists to localize proteins at a subcellular level. The analysis of the FP-tagging localization dataset along with the subcellular proteomics data, both available in SUBA3, has revealed subcellular compartments where up to 88% the FP localizations have been confirmed by subcellular proteomics for proteins for which both data are available. Thus, here the protein’s targeting ability agrees with its observed protein’s accumulation. The more data become available in the future, the better the coverage of each subcellular proteome and the higher the agreement between different methods is likely to be. However, with more data the number of disagreements between methods will also increase. Examining the number of existing disagreements between FP tagging and MS for the individual subcellular compartments has already exposed discrepancies in location attributions between the two methods as high as 39% of the total FP datasets for proteins for which both FP and MS data are available. Such a high discrepancy highlights problems with both the MS and FP tagging approaches, which are evident

when looking closely at the organelle examples of the plastid, mitochondrion and peroxisome. Apart from the technical issues and limitations of both approaches, the disagreements can also be due to unknown biology (dual-targeted proteins or dynamic proteins). Similarly, investigating the localization disagreements within the FP tagging method showed that the majority of multiple localization claims (80%) are due to multi-targeted proteins. The remaining 20% demonstrate a conflict in location attributions by different research groups and are possibly due to problems with the FP tagging approach, but may in some cases include dynamic proteins or unknown dual-targeted proteins. To be able to assess such localization data and draw conclusions about the reliability of localization methods and expose their limitations, collation of published results in databases like SUBA3 is extremely helpful. The intersections where existing

data disagree could be avenues for new biological discoveries to be made.

## ACKNOWLEDGMENTS

This work was supported by the Australian Research Council [CE0561495 to A. Harvey Millar and Ian D. Small, FT110100242 to A. Harvey Millar, DE120100307 to Sandra K. Tanz]; and the Government of Western Australia through funding for the WA Centre of Excellence for Computational Systems Biology [DIR WA CoE].

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: [http://www.frontiersin.org/Plant\\_Proteomics/10.3389/fpls.2013.00214/abstract](http://www.frontiersin.org/Plant_Proteomics/10.3389/fpls.2013.00214/abstract)

## REFERENCES

- Asano, T., Yoshioka, Y., Kurei, S., Sakamoto, W., and Machida, Y. (2004). A mutation of the CRUMPLED LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects the pattern of cell division, cell differentiation, and plastid division in *Arabidopsis*. *Plant J.* 38, 448–459. doi: 10.1111/j.1365-313X.2004.02057.x
- Benkova, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertova, D., Jurgens, G., et al. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* 115, 591–602. doi: 10.1016/S0092-8674(03)00924-3
- Berg, R. H., and Beachy, R. N. (2008). Fluorescent protein applications in plants. *Method Cell Biol.* 85, 153–177. doi: 10.1016/S0091-679X(08)85008-X
- Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010). YLoc—an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.* 38, W497–W502. doi: 10.1093/nar/gkq477
- Brugiere, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., et al. (2004). The hydrophobic proteome of mitochondrial membranes from *Arabidopsis* cell suspensions. *Phytochemistry* 65, 1693–1707. doi: 10.1016/j.phytochem.2004.03.028
- Carrie, C., Kuhn, K., Murcha, M. W., Duncan, O., Small, I. D., O'Toole, N., et al. (2009). Approaches to defining dual-targeted proteins in *Arabidopsis*. *Plant J.* 57, 1128–1139. doi: 10.1111/j.1365-313X.2008.03745.x
- Carrie, C., Murcha, M. W., Kuehn, K., Duncan, O., Barthelet, M., Smith, P. M., et al. (2008). Type II NAD(P)H dehydrogenases are targeted to mitochondria and chloroplasts or peroxisomes in *Arabidopsis thaliana*. *FEBS Lett.* 582, 3073–3079. doi: 10.1016/j.febslet.2008.07.061
- Carrie, C., and Small, I. (2012). A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim. Biophys. Acta.* 1833, 253–259. doi: 10.1016/j.bbamcr.2012.05.029
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805. doi: 10.1126/science.8303295
- Chartrand, P., Meng, X. H., Singer, R. H., and Long, R. M. (1999). Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle *in vivo*. *Curr. Biol.* 9, 333–336. doi: 10.1016/S0960-9822(99)80144-4
- Chew, O., Lister, R., Qbadou, S., Heazlewood, J. L., Soll, J., Schleiff, E., et al. (2004). A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. *FEBS Lett.* 557, 109–114. doi: 10.1016/S0014-5793(03)01457-1
- Chew, O., Rudhe, C., Glaser, E., and Whelan, J. (2003). Characterization of the targeting signal of dual-targeted pea glutathione reductase. *Plant Mol. Biol.* 53, 341–356. doi: 10.1023/B:PLAN.000006939.87660.4f
- Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE* 5:e11335. doi: 10.1371/journal.pone.0011335
- Cutler, S. R., Ehrhardt, D. W., Griffiths, J. S., and Somerville, C. R. (2000). Random GFP: cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3718–3723. doi: 10.1073/pnas.97.7.3718
- Deblasio, S. L., Sylvester, A. W., and Jackson, D. (2010). Illuminating plant biology: using fluorescent proteins for high-throughput analysis of protein localization and function in plants. *Brief. Funct. Genom.* 9, 129–138. doi: 10.1093/bfpg/elp060
- Dixit, R., Cyr, R., and Gilroy, S. (2006). Using intrinsically fluorescent proteins for plant cell imaging. *Plant J.* 45, 599–615. doi: 10.1111/j.1365-313X.2006.02658.x
- Duchene, A. M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N. M., et al. (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16484–16489. doi: 10.1073/pnas.0504682102
- Ehrhardt, D. (2003). GFP technology for live cell imaging. *Curr. Opin. Plant Biol.* 6, 622–628. doi: 10.1016/j.pbi.2003.09.014
- Escobar, N. M., Haupt, S., Thow, G., Boevink, P., Chapman, S., and Oparka, K. (2003). High-throughput viral expression of cDNA-green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. *Plant Cell.* 15, 1507–1523. doi: 10.1105/tpc.013284
- Eubel, H., Meyer, E. H., Taylor, N. L., Bussell, J. D., O'Toole, N., Heazlewood, J. L., et al. (2008). Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of *Arabidopsis* cell culture peroxisomes. *Plant Physiol.* 148, 1809–1829. doi: 10.1104/pp.108.129999
- Ferro, M., Brugiere, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., et al. (2010). AT\_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell Proteome.* 9, 1063–1084. doi: 10.1074/mcp.M900325-MCP200
- Fricker, M., Runions, J., and Moore, I. (2006). Quantitative fluorescence microscopy: from art to science. *Annu. Rev. Plant Biol.* 57, 79–107. doi: 10.1146/annurev.arplant.57.032905.105239
- Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., et al. (2004). In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* 16, 478–499. doi: 10.1105/tpc.017814
- Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., et al. (2003). Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res.* 2, 413–425. doi: 10.1021/pr034025j
- Gao, H., Kadirjan-Kalbach, D., Froehlich, J. E., and Osteryoung, K. W. (2003). ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4328–4333. doi: 10.1073/pnas.0530206100
- Gardiner, J. C., Taylor, N. G., and Turner, S. R. (2003). Control of cellulose synthase complex localization in developing xylem. *Plant*

- Cell* 15, 1740–1748. doi: 10.1105/tpc.012815
- Giacomelli, L., Rudella, A., and Van Wijk, K. J. (2006). High light response of the thylakoid proteome in Arabidopsis wild type and the ascorbate-deficient mutant vtc2-2. A comparative proteomics study. *Plant Physiol.* 141, 685–701. doi: 10.1104/pp.106.080150
- Granlund, I., Kieselbach, T., Alm, R., Schroder, W. P., and Emanuelsson, C. (2011). Clustering of MS spectra for improved protein identification rate and screening for protein variants and modifications by MALDI-MS/MS. *J. Proteom.* 74, 1190–1200. doi: 10.1016/j.jprot.2011.04.008
- Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giege, P. (2011). An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* 23, 730–740. doi: 10.1105/tpc.110.081638
- Hanson, M. R., and Kohler, R. H. (2001). GFP imaging: methodology and application to investigate cellular compartmentation in plants. *J. Exp. Bot.* 52, 529–539. doi: 10.1093/jxb/52.3.529
- Haseloff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2122–2127. doi: 10.1073/pnas.94.6.2122
- Heazlewood, J. L., Tonti-Filippini, J., Verboom, R. E., and Millar, A. H. (2005). Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. *Plant Physiol.* 139, 598–609. doi: 10.1104/pp.105.065532
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007). SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res.* 35, D213–D218. doi: 10.1093/nar/gkl863
- Hjelmstad, R. H., and Bell, R. M. (1990). The sn-1,2-diaclylglycerol cholinephosphotransferase of *Saccharomyces cerevisiae*. Nucleotide sequence, transcriptional mapping, and gene product analysis of the CPT1 gene. *J. Biol. Chem.* 265, 1755–1764.
- Hooks, K. B., Turner, J. E., Graham, I. A., Runions, J., and Hooks, M. A. (2012). GFP-tagging of Arabidopsis acyl-activating enzymes raises the issue of peroxisome-chloroplast import competition versus dual localization. *J. Plant Physiol.* 169, 1631–1638. doi: 10.1016/j.jplph.2012.05.026
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., et al. (2003). Global analysis of protein localization in budding yeast. *Nature* 425, 686–691. doi: 10.1038/nature02026
- Ito, J., Batth, T. S., Petzold, C. J., Redding-Johanson, A. M., Mukhopadhyay, A., Verboom, R., et al. (2011). Analysis of the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J. Proteome Res.* 10, 1571–1582. doi: 10.1021/pr1009433
- Jaquinod, M., Villiers, F., Kieffer-Jaquinod, S., Hugouvieux, V., Bruley, C., Garin, J., et al. (2007). A proteomics dissection of Arabidopsis thaliana vacuoles isolated from cell culture. *Mol. Cell Proteom.* 6, 394–412. doi: 10.1074/mcp.M600250-MCP200
- Kim, J. Y., Yuan, Z., and Jackson, D. (2003). Developmental regulation and significance of KNOX protein trafficking in Arabidopsis. *Development* 130, 4351–4362. doi: 10.1242/dev.00618
- Kleffmann, T., Russenberger, D., Von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., et al. (2004). The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* 14, 354–362. doi: 10.1016/j.cub.2004.02.039
- Klodmann, J., Senkler, M., Rode, C., and Braun, H. P. (2011). Defining the protein complex proteome of plant mitochondria. *Plant Physiol.* 157, 587–598. doi: 10.1104/pp.111.182352
- Koroleva, O. A., Tomlinson, M. L., Leader, D., Shaw, P., and Doonan, J. H. (2005). High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. *Plant J.* 41, 162–174. doi: 10.1111/j.1365-313X.2004.02281.x
- Lazarowitz, S. G., and Beachy, R. N. (1999). Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *Plant Cell* 11, 535–548.
- Lee, C. P., Eubel, H., O'Toole, N., and Millar, A. H. (2011). Combining proteomics of root and shoot mitochondria and transcript analysis to define constitutive and variable components in plant mitochondria. *Phytochemistry* 72, 1092–1108. doi: 10.1016/j.phytochem.2010.12.004
- Lundquist, P. K., Poliakov, A., Bhuiyan, N. H., Zybailov, B., Sun, Q., and Van Wijk, K. J. (2012). The functional network of the Arabidopsis plastoglobule proteome based on quantitative proteomics and genome-wide coexpression analysis. *Plant Physiol.* 158, 1172–1192. doi: 10.1104/pp.111.193144
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089–2103. doi: 10.1105/tpc.104.022236
- Mano, S., Nakamori, C., Kondo, M., Hayashi, M., and Nishimura, M. (2004). An Arabidopsis dynamin-related protein, DRP3A, controls both peroxisomal and mitochondrial division. *Plant J.* 38, 487–498. doi: 10.1111/j.1365-313X.2004.02063.x
- Marion, J., Bach, L., Bellec, Y., Meyer, C., Gissot, L., and Faure, J. D. (2008). Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of Arabidopsis seedlings. *Plant J.* 56, 169–179. doi: 10.1111/j.1365-313X.2008.03596.x
- Meyer, E. H., Taylor, N. L., and Millar, A. H. (2008). Resolving and identifying protein components of plant mitochondrial respiratory complexes using three dimensions of gel electrophoresis. *J. Proteome Res.* 7, 786–794. doi: 10.1021/pr700595p
- Miao, Y., and Jiang, L. (2007). Transient expression of fluorescent fusion proteins in protoplasts of suspension cultured cells. *Nat. Protoc.* 2, 2348–2353. doi: 10.1038/nprot.2007.360
- Michalecka, A. M., Svensson, A. S., Johansson, F. I., Agius, S. C., Johansson, U., Brennicke, A., et al. (2003). Arabidopsis genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light. *Plant Physiol.* 133, 642–652. doi: 10.1104/pp.103.024208
- Millar, A. H., Carrie, C., Pogson, B., and Whelan, J. (2009). Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell* 21, 1625–1631. doi: 10.1105/tpc.109.066019
- Mireau, H., Lancelin, D., and Small, I. D. (1996). The same Arabidopsis gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. *Plant Cell* 8, 1027–1039.
- Mitra, S. K., Walters, B. T., Clouse, S. D., and Goshe, M. B. (2009). An efficient organic solvent based extraction method for the proteomic analysis of Arabidopsis plasma membranes. *J. Proteome Res.* 8, 2752–2767. doi: 10.1021/pr801044y
- Mitschke, J., Fuss, J., Blum, T., Högglund, A., Reski, R., Kohlbacher, O., et al. (2009). Prediction of dual protein targeting to plant organelles. *New Phytol.* 183, 224–235. doi: 10.1111/j.1469-8137.2009.02832.x
- Murcha, M. W., Elhafez, D., Lister, R., Tonti-Filippini, J., Baumgartner, M., Philippark, K., et al. (2007). Characterization of the preprotein and amino acid transporter gene family in Arabidopsis. *Plant Physiol.* 143, 199–212. doi: 10.1104/pp.106.090688
- Nelson, B. K., Cai, X., and Nebenfuhr, A. (2007). A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* 51, 1126–1136. doi: 10.1111/j.1365-313X.2007.03212.x
- Olinares, P. D., Ponnala, L., and Van Wijk, K. J. (2010). Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell Proteom.* 9, 1594–1615. doi: 10.1074/mcp.M000038-MCP201
- Palmieri, F., Rieder, B., Ventrella, A., Blanco, E., Do, P. T., Nunes-Nesi, A., et al. (2009). Molecular identification and functional characterization of Arabidopsis thaliana mitochondrial and chloroplastic NAD+ carrier proteins. *J. Biol. Chem.* 284, 31249–31259. doi: 10.1074/jbc.M109.041830
- Parry, M. A., Andralojc, P. J., Mitchell, R. A., Madgwick, P. J., and Keys, A. J. (2003). Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exp. Bot.* 54, 1321–1333. doi: 10.1093/jxb/erg141
- Peeters, N., and Small, I. (2001). Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* 1541, 54–63. doi: 10.1016/S0167-4889(01)00146-X
- Peltier, J. B., Cai, Y., Sun, Q., Zabrowskov, V., Giacomelli, L., Rudella, A., et al. (2006). The oligomeric stromal proteome of



- Arabidopsis thaliana* chloroplasts. *Mol. Cell Proteom.* 5, 114–133. doi: 10.1074/mcp.M500180-MCP200
- Peltier, J. B., Ytterberg, A. J., Sun, Q., and Van Wijk, K. J. (2004). New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast, and versatile fractionation strategy. *J. Biol. Chem.* 279, 49367–49383. doi: 10.1074/jbc.M406763200
- Reumann, S. (2011). Toward a definition of the complete proteome of plant peroxisomes: where experimental proteomics must be complemented by bioinformatics. *Proteomics* 11, 1764–1779. doi: 10.1002/pmic.201000681
- Reumann, S., Babujee, L., Ma, C., Wienkoop, S., Siemsen, T., Antonicelli, G. E., et al. (2007). Proteome analysis of *Arabidopsis* leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. *Plant Cell* 19, 3170–3193. doi: 10.1105/tpc.107.050989
- Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., et al. (2009). In-depth proteome analysis of *Arabidopsis* leaf peroxisomes combined with *in vivo* subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. *Plant Physiol.* 150, 125–143. doi: 10.1104/pp.109.137703
- Richly, E., and Leister, D. (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. *Gene* 329, 11–16. doi: 10.1016/j.gene.2004.01.008
- Rouwendal, G. J., Mendes, O., Wolbert, E. J., and Douwe De Boer, A. (1997). Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. *Plant Mol. Biol.* 33, 989–999. doi: 10.1023/A:1005740823703
- Rudhe, C., Chew, O., Whelan, J., and Glaser, E. (2002). A novel *in vitro* system for simultaneous import of precursor proteins into mitochondria and chloroplasts. *Plant J.* 30, 213–220. doi: 10.1046/j.1365-313X.2002.01280.x
- Rutschow, H., Ytterberg, A. J., Friso, G., Nilsson, R., and Van Wijk, K. J. (2008). Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in *Arabidopsis*. *Plant Physiol.* 148, 156–175. doi: 10.1104/pp.108.124545
- Sedbrook, J. C., Carroll, K. L., Hung, K. F., Masson, P. H., and Somerville, C. R. (2002). The *Arabidopsis* SKU5 gene encodes an extracellular glycosyl phosphatidylinositol-anchored glycoprotein involved in directional root growth. *Plant Cell* 14, 1635–1648. doi: 10.1105/tpc.002360
- Skalitzky, C. A., Martin, J. R., Harwood, J. H., Beirne, J. J., Adamczyk, B. J., Heck, G. R., et al. (2011). Plastids contain a second sec translocase system with essential functions. *Plant Physiol.* 155, 354–369. doi: 10.1104/pp.110.166546
- Small, I., Wintz, H., Akashi, K., and Mireau, H. (1998). Two birds with one stone: genes that encode products targeted to two or more compartments. *Plant Mol. Biol.* 38, 265–277. doi: 10.1023/A:1006081903354
- Souciet, G., Menand, B., Ovesna, J., Cosset, A., Dietrich, A., and Wintz, H. (1999). Characterization of two bifunctional *Arabidopsis thaliana* genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. *Eur. J. Biochem.* 266, 848–854. doi: 10.1046/j.1432-1327.1999.00922.x
- Sun, X., Fu, T., Chen, N., Guo, J., Ma, J., Zou, M., et al. (2010). The stromal chloroplast Deg7 protease participates in the repair of photosystem II after photoinhibition in *Arabidopsis*. *Plant Physiol.* 152, 1263–1273. doi: 10.1104/pp.109.150722
- Tanz, S. K., Castleden, I., Hooper, C. M., Vacher, M., Small, I., and Millar, H. A. (2013). SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res.* 41, D1185–D1191. doi: 10.1093/nar/gks1151
- Tanz, S. K., and Small, I. (2011). In silico methods for identifying organellar and suborganellar targeting peptides in *Arabidopsis chloroplast* proteins and for predicting the topology of membrane proteins. *Method Mol. Biol.* 774, 243–280. doi: 10.1007/978-1-61779-234-2\_16
- Taylor, N. L., Heazlewood, J. L., and Millar, A. H. (2011). The *Arabidopsis thaliana* 2-D gel mitochondrial proteome: refining the value of reference maps for assessing protein abundance, contaminants and post-translational modifications. *Proteomics* 11, 1720–1733. doi: 10.1002/pmic.201000620
- Tian, G. W., Mohanty, A., Chary, S. N., Li, S., Paap, B., Drakakaki, G., et al. (2004). High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* 135, 25–38. doi: 10.1104/pp.104.040139
- Tsuda, K., Qi, Y., Nguyen Le, V., Bethke, G., Tsuda, Y., Glazebrook, J., and Katagiri, F. (2012). An efficient *Agrobacterium*-mediated transient transformation of *Arabidopsis*. *Plant J.* 69, 713–719. doi: 10.1111/j.1365-313X.2011.04819.x
- Ueki, S., and Citovsky, V. (2011). To gate, or not to gate: regulatory mechanisms for intercellular protein transport and virus movement in plants. *Mol. Plant* 4, 782–793. doi: 10.1093/mp/ssr060
- Yang, Y., Li, R., and Qi, M. (2000). *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J.* 22, 543–551. doi: 10.1046/j.1365-313x.2000.00760.x
- Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199
- Yu, F., Liu, X., Alsheikh, M., Park, S., and Rodermel, S. (2008). Mutations in SUPPRESSOR OF VARIATION1, a factor required for normal chloroplast translation, suppress var2-mediated leaf variegation in *Arabidopsis*. *Plant Cell* 20, 1786–1804. doi: 10.1105/tpc.107.054965
- Zacharias, D. A., and Tsien, R. Y. (2006). Molecular biology and mutation of green fluorescent protein. *Method Biochem. Anal.* 47, 83–120.
- Zhang, Z. J., and Peck, S. C. (2011). Simplified enrichment of plasma membrane proteins for proteomic analyses in *Arabidopsis thaliana*. *Proteomics* 11, 1780–1788. doi: 10.1002/pmic.201000648
- Zhou, R., Benavente, L. M., Stepanova, A. N., and Alonso, J. M. (2011). A recombineering-based gene tagging system for *Arabidopsis*. *Plant J.* 66, 712–723. doi: 10.1111/j.1365-313X.2011.04524.x
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., et al. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3:e1994. doi: 10.1371/journal.pone.0001994

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 April 2013; paper pending published: 27 April 2013; accepted: 05 June 2013; published online: 24 June 2013.

Citation: Tanz SK, Castleden I, Small ID and Millar AH (2013) Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Front. Plant Sci.* 4:214. doi: 10.3389/fpls.2013.00214

This article was submitted to *Frontiers in Plant Proteomics*, a specialty of *Frontiers in Plant Science*.

Copyright © 2013 Tanz, Castleden, Small and Millar. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.