Published in final edited form as:

Methods Enzymol. 2011 ; 491: 235–260. doi:10.1016/B978-0-12-385928-0.00014-6.

Decreased Secretion and Unfolded Protein Response Upregulation

Carissa L. Young^{*}, Theresa Yuraszeck[†], and Anne S. Robinson^{*}

^{*} Department of Chemical Engineering, University of Delaware, Newark, Delaware, USA

[†] Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California, USA

Abstract

Recombinant antibody fragments, for example, the classic monovalent single-chain antibody (scFv), are emerging as credible alternatives to monoclonal antibody (mAb) products. scFv fragments maintain a diverse range of potential applications in biotechnology and can be implemented as powerful therapeutic and diagnostic agents. As such, a variety of hosts have been used to produce antibody fragments resulting in varying degrees of success. Yeast, *Saccharomyces cerevisiae*, is an attractive host due to quality control mechanisms of the secretory pathway that ensure secreted proteins are properly folded. However, the expression of a recombinant protein in yeast is not trivial; neither are the quality control mechanisms the cell initiates to respond to overwhelming stress, such as an increased protein load, simplistic. The endoplasmic reticulum (ER) is a dynamic organelle, capable of sensing and adjusting its folding capacity in response to increased demand. When protein abundance or terminally misfolded proteins overwhelm the ER's capacity, the unfolded protein response (UPR) is activated. In the guidelines presented here, we discuss varying aspects of quality control, its modulation, and ways to design appropriate constructs for yeast recombinant protein expression. Furthermore, we have provided protocols and methods to monitor intracellular protein expression and trafficking as well as evaluation of the UPR, with essential controls. The latter part of this chapter will review considerations for the experimental design of microarray and quantitative polymerase chain reaction (q-PCR) techniques while suggesting appropriate means of data analysis.

1. Introduction

Efficient production of heterologous proteins by eukaryotic hosts, including the yeast *Saccharomyces cerevisiae*, is often hampered by the inefficient secretion of the product. Limitation of protein secretion has often been attributed to a decreased folding rate, leading to novel solutions that have implemented the overexpression of endogenous proteins to support protein folding and maturation (Robinson *et al.*, 1994; Shusta *et al.*, 1998), or the constitutive expression of the unfolded protein response (UPR) transcription factor (TF), Hac1 (Valkonen *et al.*, 2003). Extensive studies conducted with single-chain antibody fragments (scFv) established that overexpression of BiP or PDI increases secretion substantially; and when co-overexpressed, these endoplasmic reticulum (ER)-resident folding assistants act synergistically with the tuning of scFv copy number to increase secretion eightfold (i.e., 20 mg/L; Shusta *et al.*, 1998). Furthermore, our studies have shown

that overexpression of PDI or co-overexpression of BiP and PDI significantly reduce the UPR during scFv expression as determined by the canonical UPRE sensor (Mori *et al.*, 1998; Travers *et al.*, 2000). A reduction in the UPR led to an ~3.5-fold increase of secreted scFv (Xu *et al.*, 2005).

The comprehensive characterization of genes required for endogenous protein folding (Jonikas *et al.*, 2009), translational behavior (Arava *et al.*, 2003), and cell growth in *S. cerevisiae* (Brauer *et al.*, 2008) have been described, and are useful for categorizing genes of interest. Therefore, assuming that other components of the secretory pathway affect secretion although are not directly involved in protein folding, one aim of our studies was to evaluate the transcriptome of *S. cerevisiae* during expression of scFv. On a cellular level, we were also interested in understanding the interactions between BiP and PDI with the scFv during expression. We have determined UPR modulation in different cell strains and evaluated the intracellular retention of scFv based on pulse-chase ³⁵S studies to examine trafficking effects.

This chapter focuses on quality control mechanisms of the secretory pathway, specifically impacting UPR upregulation and decreased secretion of scFv. In this regard, we detail experimental methods used to evaluate the UPR in a population, and appropriate means of quantifying the intracellular concentration of a model antibody fragment, scFv 4-4-20, that may be broadly applied to heterologous protein expression and secretion. Rigorous statistical analysis of microarray and quantitative PCR (q-PCR) data is essential when evaluating global data using either a time-course or static experiment. We have carefully outlined methods and caveats in data analysis and interpretation, and utilize our studies of UPR induction by chemical treatment and expression of scFv as case studies.

2. Heterologous Protein Expression

Collectively, heterologous protein secretion involves the coupled processes of protein synthesis, protein folding, and secretory trafficking; thus, a more complete understanding of how these processes interrelate will lead to optimized conditions for scFv expression, secretion, and enhanced activity. In the case of scFv production, there are several reports in literature describing approaches to improve expression: overexpression of folding assistants BiP and PDI (Robinson *et al.*, 1994; Xu *et al.*, 2005) combined with expression level tuning of the heterologous protein (Shusta *et al.*, 1998); modulation of quality control mechanisms including the UPR (Rakestraw and Wittrup, 2006); deletions of vacuolar sorting proteins to improve secretion of heterologous proteins (Zhang *et al.*, 2001); engineered native secretory leader sequences by directed evolution to enhance scFv secretion (Rakestraw *et al.*, 2009); isolation of antibodies for increased stability and affinity using high-throughput yeast surface display (Chao *et al.*, 2006) combined with yeast cDNA overexpression libraries to identify optimal strains leading to scFv secretion (Wentz and Shusta, 2007); and mutational studies to increase the affinity of scFv (Boder *et al.*, 2000; Midelfort and Wittrup, 2006).

2.1. scFv 4-4-20 as a model scFv

To study the effects of scFv expression and trafficking on the yeast UPR, we have focused on a well-characterized construct, scFv 4-4-20, which binds to the ligand fluorescein. This

scFv quenches the fluorescence of fluorescein greater than 90% upon binding. Thus, a quick and easy spectrofluorimetric assay is used for the detection of proper folding and activity (Denzin and Voss, 1992). Expression of scFv 4-4-20 in yeast is limited by the activation of quality control machinery, and the subsequent decrease in protein production and secretion (Kauffman *et al.*, 2002).

3. Quality Control Mechanisms of the Secretory Pathway

The secretory pathway of eukaryotic cells is composed primarily of two organelles, the ER and Golgi, responsible for maintaining the fidelity of protein synthesis and maturation. The environment of the ER is specialized to properly fold secretory proteins due to an oxidizing redox potential, appropriate calcium levels, and dedicated enzymes for protein glycosylation and folding (i.e., chaperones and foldases; van Anken and Braakman, 2005). When abnormalities do occur, such as an overwhelming abundance of improperly folded protein retained in the ER, or a decrease in vesicle trafficking from the ER to Golgi, these phenomena, collectively termed “ER stress,” upregulate quality control mechanisms to ensure cellular homeostasis. Such stress can be caused by a variety of insults, including nutrient deprivation, pathogenic infection, chemical treatment, and the expression of heterologous protein.

ER-associated degradation (ERAD) and the UPR are two quality control mechanisms that are upregulated upon ER stress. They have several outcomes that occur at various timescales, depend on variations in the spatial organization of organelles, and alter selective protein concentrations and intracellular localization (Brodsky, 2007; McCracken and Brodsky, 2003; Midelfort and Wittrup, 2006; Ng *et al.*, 2000). Removal of misfolded proteins through ERAD occurs by ubiquitination via ER-associated ubiquitin-conjugating enzymes, followed by retrotranslocation, and degradation in the cytoplasm by the proteasome (reviewed in Nakatsukasa and Brodsky, 2008). Local perturbations of unfolded protein levels in the ER activate the UPR, defined as a global cytoprotective signaling cascade that transcriptionally upregulates genes encoding ERAD components, chaperones, and oxidoreductases (Otte and Barlowe, 2004; Travers *et al.*, 2000).

Once the ER's capacity to process nascent proteins is overwhelmed, the ER-resident chaperone BiP is sequestered from the transmembrane kinase protein Ire1p to counteract the stressful condition, which triggers the negative feedback loop known as the UPR. Unbound Ire1p dimerizes, autophosphorylates, and splices an intron from *HAC1* mRNA. The resulting Hac1p transcription factor (TF) binds to the promoter regions of UPR targets, upregulating their expression. However, it must also be noted that unfolded protein may directly initiate the dimerization and activation of Ire1p (Kimata *et al.*, 2007; Oikawa *et al.*, 2007). Either mechanism results in the upregulation of diverse targets that include chaperones and foldases and many elements of the secretory pathway, as well as components of the ERAD machinery. In higher eukaryotes, several additional pathways are activated in response to ER stress, and a general attenuation of translation is observed.

4. Endoplasmic Reticulum Export and Trafficking

Secreted heterologous proteins, such as scFv, enter the secretory pathway at the ER and move via vesicular transport to the Golgi. Export from the ER requires specialized ER-resident proteins including COPII machinery, and receptors such as Erv29 help export certain soluble cargo proteins (Otte and Barlowe, 2004). There is increasing evidence that post-ER quality control mechanisms exist and sort defective proteins from the Golgi to the endosomal system for lysosomal/vacuolar degradation (reviewed in Arvan *et al.*, 2002). Interestingly, *vacuolar protein sorting* (*vps*) mutants have resulted in the increased secretion of luminal recombinant proteins (Hong *et al.*, 1996) and membrane proteins (Luo and Chang, 2000).

5. Experimental Systems to Evaluate Expression, UPR, and Secretory Processing

As a simpler eukaryotic organism, yeast is advantageous to study due to facile genetic manipulation in a sequenced, annotated genome; mammalian-like organelle trafficking of secreted protein with similar mechanisms for protein synthesis, maturation, and secretory trafficking; and microbial growth features. In the following sections, we outline pertinent factors to consider with respect to molecular engineering of constructs for heterologous protein expression and selection of the most appropriate *S. cerevisiae* strain. We also outline experimental protocols and conclude with additional remarks regarding experiments and data analysis.

6. *S. cerevisiae* Strains Used for Optimal Expression

A yeast strain should be selected based on its suitability for the process being studied, efficiency of transformation, and flexibility with respect to selection. Difficulties associated with the expression level of a recombinant protein, effect of growth rates, and proteases are aspects that should be considered. The choice of an appropriate host strain, induction media, and expression plasmid (i.e., 2 μ m, low-copy, or multicopy δ integrating plasmids) can overcome most obstacles. Usually it is desirable to choose a specific parental strain that has been used in previous studies (or industrial applications), therefore allowing direct comparison with established results and not complicating your analysis by differences in strain backgrounds. Additionally, consider strains that carry multiple deletion alleles of auxotrophic markers that will provide flexibility in the future should you choose to introduce episomal plasmids or PCR-based modifications completed by homologous recombination (Brachmann *et al.*, 1998). Deletions of nearly all ~ 6000 genes have been conducted in multiple *S. cerevisiae* strains (see yeast gene knockout or YKO Collection, Open Biosystems) providing remarkable options. Alternatively, it is rather straightforward to design additional auxotrophic knockouts in your strain of choice (Petracek and Longtine, 2002).

To alleviate the problem of contaminating proteases, a protease-deficient strain (BJ5464 MATa ura3-52 trp1 leu2 his3 200 pep4::HIS3 prb1- 1.6R can1 GAL (ATCC 208288)), including mutations in both the *pep4* and *prb1* genes, is recommended (reviewed by Jones,

2002). However, one must keep in mind that all vacuolar proteases increase in concentration as the cells approach stationary phase, and a small increase has been observed at the diauxic plateau; the largest fold increase (i.e., 100× that of log phase) occurs as the cells enter stationary phase (Moehle *et al.*, 1987). Therefore, it is imperative that strain growth and heterologously expressed proteins are monitored by time-course studies. An orthogonal challenge is that pep4 and prb1 are specific precursors for degradation of autophagosomes; therefore, the use of these mutants will potentially lead to an accumulation of terminally misfolded proteins in the cell that may become detrimental to cell growth and homeostasis (Takeshige *et al.*, 1992).

7. Plasmid Design

For the maximal expression of a recombinant protein, the 2 μ m is generally considered. Yet, high-level expression will overwhelm the capacity of the secretory pathway; therefore, alternatives include low-copy plasmids (i.e., pRS300 series; Sikorski and Hieter, 1989), direct chromosomal integration, or the flexibility provided by δ integration. Comparison of these molecular engineered constructs are examined in Parekh *et al.* (1996), Shusta *et al.* (1998), and Kauffman *et al.* (2002) by evaluating heterologous protein secretion. When deciding on the appropriate vector, keep in mind that the use of minimal media is essential for auxotrophic markers, resulting in a doubling time of approximately 140 min. This limitation can be avoided by the integration of a PCR-based fragment resulting in a doubling time of 90 min in rich media (Sherman, 2002).

Promoter choices range from constitutive promoters (e.g., PGK and GAPDH) to inducible promoters (e.g., GAL1, GAL1-10, CUP1). Galactose-regulated promoters are induced >1000-fold, yet are strongly repressed by glucose; thus, maximal expression only occurs following glucose depletion (Romanos *et al.*, 2001). For rapid induction, cultures can be grown overnight using a nonrepressing carbon source (e.g., raffinose) and then induced by direct addition of galactose (review protocol in Romanos *et al.*, 2001). Furthermore, regulation of gene expression can be conducted by the tetracycline/tetracycline operator (TetO) systems (Belli *et al.*, 1998) or use of the ADH1 promoter that eliminates the need to shift carbon sources that can alter metabolism and gene expression (Gasch *et al.*, 2000). Expression of scFv in yeast, *S. cerevisiae*, has utilized the GAL1, GAL1-10, and GAPDH promoters (Huang *et al.*, 2008; Shusta *et al.*, 1998; Xu and Robinson, 2009).

7.1. scFv 4-4-20 construct

The autonomous (CEN/ARS) plasmid pRS316-FLAG-4-4-20 consists of an inducible GAL1-10 promoter, synthetic pre-pro sequence (Clements *et al.*, 1991), Glu-Ala-Arg-Pro spacer, FLAG (amino acid sequence Asp-Tyr-Asp-Asp-Asp-Lys) epitope tag, scFv gene, and α -terminator sequence that has been used for pulse-chase studies (Figs. 14.1 and 14.3). Site-directed mutagenesis of two initial stop codons following the scFv 4-4-20 gene restored the original C terminal c-myc epitope (Xu *et al.*, 2005) modified from the construct of Shusta *et al.* (1998). This low-copy plasmid was used in our strain-dependent studies, described in Fig. 14.2, described below. To examine the effects of using a multicopy integration of 4-4-20, the coding sequence described above was subsequently inserted into the δ vector, pITy-4 (Parekh *et al.*, 1996), and transformed to the BJ5464 parental strain,

referred to as high-copy scFv (hcscFv; Kauffman *et al.*, 2002). Integrated hcscFv was grown in 1% raffinose, induced by addition of galactose (1%), and samples were collected at various times ($t = 0, 2, 4, 6, 8$, and 12 h). Microarray analysis of this data described later in this chapter has identified novel regulation during heterologous recombinant protein expression.

7.2. UPR sensors

The canonical UPR element (UPRE) was originally defined as a 22-bp sequence element of the KAR2/BiP promoter (Mori *et al.*, 1992) and refined to nucleotide precision as a semipalindromic seven-nucleotide consensus, CAGNGTG (Mori *et al.*, 1998), defined as the classical UPR sensor, UPRE-1. A bioinformatics approach conducted by Patil *et al.* (2004) has identified two novel UPRE promoters with the specific nomenclature of UPRE-2 and UPRE-3. In previous studies (Xu *et al.*, 2005), we subcloned a construct containing four tandem repeats of UPRE-1, a disabled *Cyc1* promoter, and green fluorescent protein (GFP) from pKT058 (Travers *et al.*, 2000) into the low-copy pRS300 plasmids of Sikorski and Hieter (1989). Subsequent constructs included four tandem repeats of UPRE-2 and UPRE-3 in the same base vector, thereby providing an appropriate means to compare all three elements and their regulation during UPR induction (Fig. 14.1). UPR elements fused to fluorescent variants have been integrated in the chromosome of several *S. cerevisiae* strains, resulting in strains capable of growth on rich media and improved fluorescence properties (Robinson Lab, unpublished).

8. Evaluation of Heterologous Protein Expression

8.1. Strain growth, expression, and isolation of intracellular heterologous protein

The following time-course protocol is specified for the expression of a heterologous protein and evaluation of the UPR (Fig. 14.2) although it can be modified for any experimental system. Synthetic media has been described elsewhere (Sherman, 2002).

1. Following the restreak of your desired strain onto a selective agar plate (e.g., SD-URA-LEU), resuspend a single colony in 1 mL of appropriate dextrose media and measure the optical density at 600 nm (OD_{600}). Estimate, *a priori*, the volume of the resuspended sample necessary for an overnight culture to achieve a final value of $0.6 \leq OD_{600} \leq 0.8$, assuming a doubling time of 140 min (synthetic media).
2. Grow the desired volume of overnight culture in minimal dextrose media supplemented with the appropriate amino acids (e.g., SD-URA-LEU) at 250 rpm, 30 °C.
3. The following morning, measure the OD_{600} . Calculate the volume to be added for the induction phase, targeting an $OD \sim 0.1\text{--}0.3$ of the desired total volume of culture.
4. Harvest the necessary cell volume by centrifugation (2500×g, 5 min, 4 °C), discard the culture supernatant, and keep the cell pellet on ice. Wash away residual glucose by resuspending the pellet in 10 mL of dH₂O. Centrifuge again and repeat.

5. Add the resulting cell pellet to the desired total volume of galactose medium (SG-URA-LEU). For each sample (i.e., desired time points) remove 1 OD₆₀₀ equivalent volume. Spin down (2500×g, 5 min, 4 °C). Discard the supernatant, resuspend in 0.5 mL dH₂O, and transfer resuspended cells into 1.5-mL eppendorf tubes. Spin down (1 min, 13,000 rcf), remove supernatant, and freeze cell pellet at –80 °C until subsequent analysis.
6. Remove cell pellets from –80 °C and keep on ice from this point forward. Add 100 µL lysis buffer (1% SDS, 50 mM Tris–HCl, pH = 7.4, complete, EDTA-free protease inhibitor cocktail tablets, per manufacturer’s instructions (Roche)) per OD₆₀₀, adapted from Rothblatt and Schekman (1989). Break the cells by adding roughly the same amount of glass beads (0.5 mm Zirconia/Silica beads, Biospec Products, Inc.) as the volume of cell pellet (100 µL equivalent). Vortex on high speed three times for 30 s each and between each cycle chill on ice for a minimum of 30 s.
7. Retrieve the cell lysate by using gel-loading tips, transferring the lysate to a fresh 1.5-mL eppendorf tube. Clarify the extract by centrifugation at top speed for 1 min. Transfer the supernatant to a new tube.
8. Normalize all sample volumes to an equivalent OD₆₀₀ for loading. For measuring intracellular scFv levels, we typically combine 20 µL total lysate with 40 µL 3× SDS loading buffer (150 mM Tris, pH = 6.8; 0.25 mg/mL bromophenol blue; 6% SDS; 30% glycerol; 30 mM dithiothreitol (DTT)). For Western blot analysis, denature samples by heating to 95 °C for 5 min, followed by quenching on ice. A quantitative analysis of the UPRE-GFP sensor can be completed by in-gel fluorescence (see protocol, Section 8.2, below); however, samples should not be heated in order to avoid loss of fluorescence through GFP unfolding.
9. Samples are analyzed by SDS-PAGE, transferred to nitrocellulose paper, and quantified by standard Western blot procedures. Use of primary and secondary antibodies should be optimized according to levels of expressed recombinant protein. See Table 14.1 for more information regarding antibodies currently used in our laboratory. Resulting Western blots are then scanned (Typhoon 9400 Variable Mode Imager, Amersham Biosciences) and quantified using ImageQuant (Molecular Dynamics) or ImageJ (National Institutes of Health) software.

8.2. Analysis of GFP fusions by in-gel fluorescence

Analysis of fusions by in-gel fluorescence was originally described in Newstead *et al.* (2007). Implementation of this technique has proven to be a rapid method to confirm the expression of endogenous and heterologous fusion proteins in *S. cerevisiae* using multiple GFP and DsRed derivatives Robinson Lab, unpublished.

When using fluorescent protein fusions to quantify protein levels, it is extremely important to normalize protein samples correctly (e.g., to equivalent OD₆₀₀) while accounting for differences in fluorescence excitation and emission spectra, as well as intensity differences (i.e., quantum yield, extinction coefficient; reviewed in Nagai *et al.*, 2002; Rizzo *et al.*,

2004; Shaner *et al.*, 2004). When quantifying total concentration based on intensity of a fluorescent variant there are a few aspects to consider, including cell preparation (i.e., culture conditions and lysis preparation), storage (e.g., -80°C or prepared directly following experiment), and appropriate controls. We also recommend addition of fluorescent standards to determine approximate protein concentration.

Sample preparation is similar to the previous protocol (steps 1–8), although samples should not be heated. Consider the effects of storing your samples at -80°C and repeated freeze-thaw cycles, which may result in a significant decrease in fluorescence intensity. Typically, we use a sample load equivalent to $\sim 0.1 \text{ OD}_{600}$ and analyze by SDS-PAGE; however, conditions should be optimized for the specific fusion protein of interest. For in-gel fluorescence, protein bands of GFP (or DsRed) variants are imaged using a Typhoon 9400 Variable Mode Imager (Amersham Biosciences) and compared to fluorescent molecular weight ladders at 488 (BenchmarkTM Fluorescent Protein Standard, Invitrogen) or 546 nm (DyLightTM 546/649 Fluorescent Protein Molecular Weight Markers, Thermo Fisher Scientific, Inc.). It may be necessary to dilute these ladders in $1\times$ SDS running buffer in order to optimize comparison to protein bands of interest.

8.3. Experimental results

The expression and intracellular retention of scFv 4-4-20 in the parental strain BJ5464, a strain overexpressing BiP herein referred to as HBiP, and a strain co-overexpressing BiP and PDI (i.e., HBiPPDI), were examined as a function of time. UPR induction by in-gel GFP fluorescence (Fig. 14.2), intracellular scFv levels via Western analysis (Fig. 14.2), and BiP/Kar2 protein levels determined by Western blot analysis (data not shown) was measured in all three strains. We confirm that the UPR is significantly upregulated in HBiP as opposed to HBiPPDI (Fig. 14.2A). Interestingly, the UPR response as determined by fluorescence is significantly greater in the parental strain expressing 4-4-20 (i.e., BJ5464); furthermore, when compared to the Western blot analysis of Fig. 14.2B, it is readily apparent that total intracellular scFv protein levels are significantly greater in this parental strain as compared to the strains overexpressing chaperone and/or foldase. In all cases, at 24 h postinduction, the immature scFv species accumulates intracellularly causing an overwhelming stress. Thus, to investigate protein synthesis of this model scFv and concurrent trafficking defects, we recommend taking advantage of the pre-pro leader sequence combined with pulse-chase techniques that will yield valuable insights into quality control mechanisms of the secretory pathway.

8.4. Pulse-chase and immunoprecipitation protocol

To monitor intracellular trafficking of newly synthesized scFv, we have constructed strains in which ER luminal chaperone and foldase, BiP and PDI, respectively, have been overexpressed in order to increase secreted yields of scFv 4-4-20 (Xu *et al.*, 2005). Cells were grown overnight in 5 mL minimal medium (SD-2XCAA; Witttrup and Benig, 1994) supplemented with the appropriate amino acids. The overnight culture was then diluted to an $\text{OD}_{600} \sim 0.1$ in fresh SD media for ~ 6 h. In order to continuously provide nutrients to the culture, semicontinuous growth was implemented once expression was induced via galactose (1%). A 0.2 OD_{600} equivalent of cells were resuspended in SG-2XCAA media

lacking uracil (SG-URA). Yeast strains were grown semicontinuously in order to maintain the culture between 0.2 OD_{600} –0.3 by supplementing SG-URA media every 2 h, as needed, for 14 h. Cells were harvested, washed with water to remove methionine, and resuspended at a density of 5 OD_{600} in prewarmed SG-URA minus methionine for 1 h. The culture was then incubated with 240 μ Ci/ OD Trans³⁵S-label (MP Biomedicals, Inc.) for 10 min in an air shaker at 30 °C. Following the pulse, excess cold methionine (100mM) was added. For each time point, 1 OD_{600} of radiolabeled cells were collected by centrifugation at 14,000 rpm for 2 min. Pellets were then resuspended and incubated in 1 mL chase media consisting of SG media plus 1 mg/mL methionine per OD_{600} .

Radiolabeled scFv samples were immunoprecipitated with cells lysed in lysis buffer (2% SDS, 90 mMHEPES, pH 7.5, 30 mMDDTT, and 0.4 mg/mL Pefabloc; Sigma) and denatured at 95 °C for 10 min. Four microliters of the primary antibody (α -FLAG M2 monoclonal, Sigma) was added to each sample for a 1-h incubation at 4 °C, followed by addition of 50 μ L of resin (Protein A/G Sepharose Mix, Amersham) and an additional hour incubation at 4 °C. Samples were then centrifuged and the pellets rinsed with cold washing buffer (50 mM Tris, pH 7.5, 250 mM NaCl, 5 mM EDTA, 0.5% NP-40, 0.4 mg/mL Pefabloc, 10 μ g/mL Leupeptin, 10 μ g/mL Pepstatin, 5 μ g/mL Aprotinin) three times prior to heating to 95 °C for 5 min. scFv samples were loaded onto a 11% SDS-PAGE gel; following electrophoresis, the gel was then submerged in Amplify (Amersham) for 30 min, dried, and exposed to a Phosphorimager Scanner SI (Molecular Dynamics). The data was then scanned (Typhoon 9400 Variable Mode Imager, Amersham Biosciences) and quantified using ImageQuant software (Molecular Dynamics).

Figure 14.3 depicts typical results obtained from this pulse-chase analysis under conditions in which the culture density was maintained between 0.2 OD_{600} –0.3 by providing a continuous source of nutrients in semibatch growth. Proper processing of the pre-pro leader sequence should result in a mature scFv species that is secreted to the media. Our results confirm the existence of a mature species in the parental strain (i.e., BJ5464) and HBiPPDI; however, the predominant intracellular scFv species in the HBiP strain is an immature product at all time points. In order to elucidate the specific regulation of protein synthesis, degradation, and ER-to-Golgi trafficking mechanisms, these studies led to further analysis via microarray studies of the global gene expression changes induced by expression of a scFv construct integrated multiple times within the chromosome (i.e., hcscFv) and comparison to the UPR resulting from chemical stress (i.e., DTT addition).

8.4.1. Additional remarks—Accurate quantification of heterologous protein expression and UPR induction requires appropriate normalization procedures. Using these protocols, we normalize to number of cells assayed (via OD_{600}) and compare, by Western blot results, the magnitude of the UPR response and scFv expression to a loading control, Act1. Most importantly, by monitoring time-course expression, we can choose the optimal growth phase to harvest cells in order to achieve reproducible results. It is important to always maintain the appropriate controls, for example, the original plasmid minus the gene of interest, the parental strain without modifications, or multiple integrations of the product compared to a low-copy plasmid or single integration. Examples of these experimental conditions for optimized scFv secretion have been detailed elsewhere (Shusta *et al.*, 1998; Xu *et al.*, 2005).

Investigating one strain under one condition with different constructs, and carrying out biological replicates, will ensure reliable results.

9. Statistical Analysis of Microarray Results

The advent of high-throughput microarray technology to simultaneously measure the expression of thousands of genes has revolutionized the study of biological systems. A single microarray experiment provides an abundance of information from which diverse analyses are possible, from large-scale characterization of the effects of the experimental condition on known cellular processes to network inference efforts in which the regulatory pathways responsible for the system's response to the imposed condition are deduced. However, the obvious benefits of such technology are accompanied by the corresponding challenge of extracting meaningful insight from data that is comprehensive in genomic coverage but generally restricted with respect to aspects such as time-course measurements and biological replication. The challenge is compounded by the lack of a standard approach to microarray data analysis; the appropriate method often depends on the particulars of the experiment that generated the data.

Here we discuss the issues associated with the analysis of microarray data in the context of experiments performed to elucidate the UPR in *S. cerevisiae* (Fig. 14.4). In our experiments, we studied the UPR as activated by chemical treatment with DTT, a well-established UPR inducer, and by heterologous protein expression of hscFv. RNA isolation methods were performed as described in Current Protocols in Molecular Biology (Ausubel *et al.*, 1987). Probe preparation and microarray hybridization were completed by the modified protocol of the UHN Microarray Centre (Canada). Arrays were scanned using ScanArray Express (Microarray Scanner, PerkinElmer) and analyzed by Quantarray software (BioChip Technologies LLC). Additionally, the intensities were normalized by the LOWESS method of TIGR MIDAS (Saeed *et al.*, 2003).

9.1. Appropriate controls

The selection of appropriate controls must be completed prior to the start of the experiment and is dependent upon the nature of the experiment. In the study of the UPR in *S. cerevisiae*, samples were removed from the cell culture at various times following induction. It is desirable, although not strictly necessary, to prepare samples from the treated and control cultures prior to initiating the treatment, at $t = 0$. In our experiment, specifically pertaining to UPR induction by DTT, we account for these possible discrepancies by dividing the original culture at $t = 0$ to separate aliquots (i.e., the negative control and treated sample (i.e., DTT addition)). Minor variation in the conditions of the cultures or even stochastic variation could lead to detectable and significant differences in gene expression; to compensate for these differences and home in on the effect of stress and the ensuing initiation of the UPR, the expression measurements should be normalized to those at $t = 0$. It is not uncommon to assume that the differences between the samples prior to treatment are negligible, but it is prudent to confirm such assumptions.

To further isolate the direct effects of UPR initiation from other influences (such as a metabolic shift), analysis of samples from *hac1*, *ire1*, and/or *hac1 ire1* mutants should

be considered. Such analysis was undertaken in the canonical work by Travers *et al.* (2000), who induced the UPR by DTT treatment. However, this approach necessarily assumes that the only means by which *S. cerevisiae* responds to ER stress is via the generally accepted IRE1/HAC1 pathway; although pathways paralleling those seen in the higher eukaryotes have not yet been found, such an approach eliminates the possibility.

9.2. Selecting differentially expressed genes

Image analysis and normalization is a fairly straightforward task, with LOW-ESS normalization typically applied to eliminate systematic bias in the data (Quackenbush, 2002). Once these tasks have been completed, differentially expressed genes are selected. Several factors must be considered to select an appropriate method, most importantly whether the study is a time-course or static experiment and the number of available replicates. For a static, two-state experiment, a very popular method is the significance analysis of microarrays (SAM), a modified version of the standard *t*-test that was developed specifically to handle microarray data (Tusher *et al.*, 2001). This method requires replicates and makes no underlying assumptions about the distribution of the data. It reports the false discovery rate as the metric for determining significant expression, a metric superior to the *p*-value.

A new mixture model method in which transformed signal intensities are characterized by a set of overlapping beta distributions is also a good choice for selecting differentially expressed genes in a two-state experiment (Ogunnaïke *et al.*, 2010). Starting from first principles, the authors showed that this distribution is an adequate descriptor of microarray data and from the set of overlapping fits they calculate the probability that a gene is upregulated, downregulated, or not regulated (Ogunnaïke *et al.*, 2010). Replicates can be used to calculate a confidence metric for the reported probabilities (Ogunnaïke *et al.*, 2010). This method has the advantages of requiring no replicates and of leveraging all the data in making decisions about the differential expression of each gene, an approach that recognizes gene expression values are not independent (Ogunnaïke *et al.*, 2010). Furthermore, it slightly outperformed SAM in a simulation study, recovering 99.6% of differentially expressed genes with five replicates, although SAM's performance is still remarkable with a recovery of 94.7% of the genes (Ogunnaïke *et al.*, 2010). A point and click MATLAB implementation is available from the author's Web site (Gelmi).

In the case of a time-course study, such as our microarray analysis of hcscFv expression in *S. cerevisiae*, a method specific to detecting changes over time should be used. Several such methods exist, including the method proposed by Storey *et al.* (2005), which explicitly accounts for the temporal ordering of the data. In this construction, one is testing whether the entire time-course profile is flat (the null hypothesis). Of course, no profile is ever truly flat, but the variation in some profiles is insignificant and not related to the experimental treatment, but rather a result of limitations of the technology and noise inherent in data collection. To test for differential expression of a given gene, two curves are fit to the expression data for that gene. The first is simply the best-fit curve, which can be fit with a higher order polynomial or cubic splines. The second fitted curve is the best flat line through the data. Then a test statistic that quantifies the relative goodness of fit of the two curves is

calculated. To determine significance, a null distribution of test statistics is constructed from the replicate data. Because replicates should, theoretically, be exactly the same value, any discrepancy between them can be attributed to the inherent noise in the experiment that is not related to the biological process under consideration. Null profiles are created from the difference between the replicates, from which the null distribution of test statistics is constructed.

Replicates are not required at every time point nor for every gene, but a sufficient number of replicate permutations must be completed to create a reasonably sized null distribution; 1000 points should be sufficient, although it is often possible to calculate a much larger distribution for little additional computational expense. The main disadvantage of this method is the generally short time courses obtained via microarrays; this method works best when at least 12 points have been gathered, due to overfitting issues. An open-source software package with a user-friendly interface has been developed to complete this differential expression analysis (Leek *et al.*, 2006).

For time-course studies with less than 12 data points, an effective alternative has been proposed (Di Camillo *et al.*, 2007). Rather than fitting curves to the data, the area between the log transformed expression profile and the null profile, a flat line centered at zero, is calculated. Null distributions are again calculated from the replicate data. It was shown that this method was more effective at reducing false positives than the curve-fitting method for shorter time series. We implemented this method in MATLAB (Mathworks, Cambridge, UK), and based on our experience, we suggest this method for time-course studies with few time points.

9.3. Classification

The large amount of data generated in a microarray experiment necessitates a means of consolidating and interpreting that data. Hence the development of structured vocabularies, such as the gene ontology (GO), which provide a means of consolidating information of genes according to the biological process in which they are involved, their molecular function, or their cellular localization (Ashburner *et al.*, 2000). The GO is organized as a directed acyclic graph and its descriptions of categories are very detailed, leading to a large number of terms. The Functional Catalog (FunCat) is an alternative to the GO and does not attempt to exhaustively characterize protein function to its most granular level, but instead classifies proteins according to functional modules (Ruepp *et al.*, 2004). The choice of databases is somewhat arbitrary, as they will provide reasonably consistent results insofar as the categories in each scheme are parallel, but the FunCat's focus on functional modules facilitates data interpretation.

Regardless of the database, two important issues must be considered in classifying interesting genes from microarray experiments. First, the background set against which the interesting set is compared must be established. In most cases, the best choice of a background set is the set of genes measured via microarray; this is not necessarily the set of genes in the genome. If the microarray is enriched in genes associated with metabolism, for example, but the genome is used for the background set, it is likely that the differentially expressed genes will be highly enriched in metabolism. Yet, this result will be a function of

the enrichment of the microarray, and not truly a result of the experimental protocol. The second important issue to consider is the statistical significance of the results. Returning to our previous example, if 50% of the genes on the microarray are associated with metabolism, and 50% of the differentially expressed genes are associated with metabolism, the result might simply be a function of random chance rather than associated with the biological phenomena under study. That is not to say that the results are not biologically valid, as our upregulated group was enriched in genes encoding proteins involved in protein folding and stabilization but the conservatively corrected p -value was above the significance threshold. This finding is supported by established knowledge about the UPR, bolstering our confidence in its relevance, but in the absence of other evidence a systematic, well-supported conclusion could not be drawn. Therefore, it is important to evaluate the statistical significance of the results, an easy task to which the hypergeometric distribution function can be applied. The results must also be corrected for the multiple comparisons made in any such classification effort. The Bonferonni correction is typical, although it is very conservative. According to this correction, our upregulated gene group was enriched most notably in genes encoding proteins residing in the ER, with a Bonferonni corrected p -value of 0.02. An alternative approach is to convert the p -values derived from the hypergeometric distribution into q -values, which are the false discovery rates that would be incurred if the associated category were called significant.

9.4. Enrichment of promoter regions in known transcription factor-binding sites

Selecting differentially expressed genes and classifying them according to their biological function provides a great deal of information about the cellular response to ER stress, but it does not illuminate the regulatory pathways by which the response occurs. To address this question, the YEASTRACT database is singularly appropriate; it contains a record of all the documented associations between known *S. cerevisiae* TFs and their target genes, based on deletion experiments, band-shift, footprinting, or ChIP assays (Teixeira *et al.*, 2006). These associations may be direct or indirect. The database also contains information about potential associations based on the presence of known TF-binding site sequences in the promoter regions of *S. cerevisiae* genes. Using sets of differentially expressed genes as inputs, all known TF associations can be mined from this database. One must be careful not to simply assume the TF with the highest number of associations is responsible for the observed phenomena. Rather, evaluations similar to those applied to the classification effort must be completed to find statistically significant associations; again, the hypergeometric distribution is appropriate for this task.

Consider the results we obtained from our microarray experiment in which DTT was used to trigger the UPR. We identified a group of upregulated genes enriched for protein fate, an expected outcome given the knowledge of UPR effects. The TF Ste12p is directly associated with 30% of the genes in this group, while the known UPR regulator Hac1p is only directly associated with 4% of the group. However, when compared against the background set, the Bonferonni corrected p -value associated with Hac1p is $2e^{-4}$, while that associated with Ste12p is 0.5. Only by considering the statistical significance was the known UPR regulator (Hac1p) recovered from this analysis of the data. Discovering new regulators requires the same rigor, lest experimental resources be wasted on unpromising leads.

9.5. Presence of UPRE-1, UPRE-2, UPRE-3 in the promoter region of differentially expressed genes

Greater confidence in new results, such as the novel downregulation that we observe in our experiments to elucidate the UPR, can be obtained by evaluating the upregulated group for consistency with known facts about the UPR. Three TF-binding sites that confer upregulation on UPR targets have been identified (Mori *et al.*, 1998; Patil *et al.*, 2004b; Fig. 14.1); therefore the upregulated group should show significant enrichment in these binding sites. To determine their enrichment, the promoter regions of the upregulated genes must be extracted, necessitating a decision about what constitutes a promoter region; 600–1000 bp upstream of the start codon is a commonly used definition. The degree of allowable mismatch and the presence of the binding sites in the background set must be determined. Then, enrichment and statistical significance can be calculated. In our experiments, we found significant enrichment for these three binding sites; 42% of the upregulated group contained one of the three sites at a 0.03 level of significance.

9.6. Relative quantification by real-time quantitative PCR

Microarray data is inherently noisy and conclusions from these experiments are based on statistical analyses, necessitating further validation that is typically accomplished with q-PCR experiments. Data collected from q-PCR experiments is still widely processed by the so-called 2^{-C_t} method. The full derivation is supplied in Livak and Schmittgen (2001), but the relevant assumption here is that the efficiency with which all RNA is amplified is 100%. However, the fold-change calculation is very sensitive to the amplification efficiency and any violations of this assumption result in significant variability in the results. Bustin and Nolan (2004) observe that the efficiency of reverse transcription is low for genes expressed at low levels and point out that this contributes to the variability and lack of reproducibility of a q-PCR experiment. Ramakers *et al.* (2003) showed that theoretically, the effect of 20% variability in efficiency can result in fold-change results spanning three orders of magnitude, making it impossible to even determine the direction in which the gene was regulated. In practice, assuming efficiencies are close to 100% can be wildly inaccurate and one can run afoul of the aforementioned problems; in repeated trials, we found that the efficiency with which a canonical UPR gene, KAR2/BiP, was amplified remained consistently at <80%.

Several methods can be used to alleviate problems associated with less-than-ideal efficiencies. The current gold standard is utilizing a standard curve, but this approach has its drawbacks (Rutledge and Stewart, 2008a). Not only is it resource intensive, it also assumes that the efficiency with which any particular sample is amplified is constant (Rutledge and Stewart, 2008a). Although the latter issue may not be significant, as it appears that the variability associated with the sample is due to random error and is relatively less important than the variability associated with the amplicon (Karlen *et al.*, 2007; Nordgård *et al.*, 2006), the former issue remains a problem.

To circumvent this excessive use of resources, methods that determine efficiency directly from the amplification curves can be used. These methods are based on curve fitting and are generally based on either sigmoidal or exponential fits. To calculate fold changes in expression for microarray validation, we suggest the “linear regression of efficiency” (LRE)

analysis proposed by Rutledge and Stewart (2008b) to calculate efficiencies for use in Eq. (14.1) which gives the fold change for gene “X.”

$$\text{Fold change} = \frac{\left\{ \frac{(1+E_{hk,treated})^{C_{T,hk,treated}}}{(1+E_{x,treated})^{C_{T,x,treated}}} \right\}}{\left\{ \frac{(1+E_{hk,control})^{C_{T,hk,control}}}{(1+E_{x,control})^{C_{T,x,control}}} \right\}} = \frac{(1+E_{hk})^{C_{T,hk,treated}-C_{T,hk,control}}}{(1+E_x)^{C_{T,x,treated}-C_{T,x,control}}} \quad (14.1)$$

where $E_{hk,treated}$, $C_{T,hk,treated}$, $E_{hk,control}$, and $C_{T,hk,control}$ are the efficiencies (E) and C values, respectively, associated with the housekeeping (hk) gene in the treated and control samples, and $E_{x,treated}$, $C_{T,x,treated}$, $E_{x,control}$, and $C_{T,x,control}$ are the efficiencies and C_T values associated with the gene “X” for which the fold change is being calculated in the treated and control sample.

Although this method was developed with absolute quantification in mind, it is easily applicable to the relative quantification of RNA. It assumes that fluorescence and amplicon quantity are proportional and has the advantage of accounting for the decline in efficiency over the course of the amplification reaction (Rutledge and Stewart, 2008b); most methods assume that the efficiency is constant in the log-linear region. Based on results showing that the variability in efficiency is random and most strongly influenced by sequence (Karlen *et al.*, 2007; Nordgård *et al.*, 2006), we suggest merging this method with that of Karlen *et al.* (2007) to provide a robust measure of efficiency. Instead of using sample-specific efficiencies, the efficiency should be taken as the average over all the samples and replicates for an amplicon; such an approach was found to be highly effective (Karlen *et al.*, 2007). Furthermore, it is important to consider the threshold at which the C_T values are calculated. It is common to choose a single value for all the amplicons studied, but this approach does not guarantee that the threshold will fall in the log-linear region of the amplification curve. Thus, it is suggested that the threshold be chosen in an amplicon-specific manner.

Dealing with replicates and establishing error estimates is a subject that is not often addressed in the methodology papers for q-PCR analysis, but is important for establishing the statistical reliability of the validation results. At least three replicates should be used to guarantee statistically significant results and possibly more, depending on the induction ratio of the amplicon being validated and the reproducibility of the amplification reaction curves (Karlen *et al.*, 2007). To estimate the error, standard error propagation formulas should be used (Nordgård *et al.*, 2006).

9.7. Implications for protein secretion and UPR induction

The application of microarray technology and appropriate analysis methods contribute to our understanding of protein secretion in *S. cerevisiae* and the effect of the UPR on cellular function. In particular, we found that four key trafficking proteins, previously unreported, were upregulated in response to UPR induction by DTT: Sec72, Sil1, Erv29, Erp2. Specifically, Sec72 is a component of the ER translocation complex, Sil1 is a nucleotide exchange factor of chaperone BiP/Kar2, and Erv29 and Erp2 are products of ER-to-Golgi trafficking localized to COPII machinery. Interestingly, Erv29 directly interacts with a

specific motif on the pro-signal sequence and is suggested to direct soluble cargo to vesicle components of the secretory pathway (Otte and Barlowe, 2004). Rigorous analysis of microarray data under various environmental stimuli combined with complementary diverse experimental methods has resulted in an improved understanding of scFv expression, trafficking, and ER quality control in yeast, *S. cerevisiae*.

ACKNOWLEDGMENT

The authors were supported in part for the experimental studies by NIH R01 GM 075297.

REFERENCES

- Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*. 2003; 100:3889–3894. [PubMed: 12660367]
- Arvan P, Zhao X, Ramos-Castaneda J, Chang A. Secretory pathway quality control operating in Golgi, plasmalemmal, and endosomal systems. *Traffic*. 2002; 3:771–780. [PubMed: 12383343]
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*. 2000; 25:25–29. [PubMed: 10802651]
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Purification and concentration of RNA and dilute solutions of DNA. *Current Protocols in Molecular Biology*. 1987 section 2.1.7 (Supplement 59).
- Belli G, Gari E, Aldea M, Herrero E. Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. *Yeast*. 1998; 14:1127–1138. [PubMed: 9778798]
- Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. USA*. 2000; 97:10701–10705. [PubMed: 10984501]
- Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*. 1998; 14:115–132. [PubMed: 9483801]
- Brauer MJ, Huttenhower C, Airoidi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol. Biol. Cell*. 2008; 19:352–367. [PubMed: 17959824]
- Brodsky JL. The protective and destructive roles played by molecular chaperones during ERAD (endoplasmic-reticulum-associated degradation). *Biochem. J*. 2007; 404:353–363. [PubMed: 17521290]
- Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech*. 2004; 15:155–166. [PubMed: 15331581]
- Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc*. 2006; 1:755–768. [PubMed: 17406305]
- Clements JM, Catlin GH, Price MJ, Edwards RM. Secretion of human epidermal growth factor from *Saccharomyces cerevisiae* using synthetic leader sequences. *Gene*. 1991; 106:267–271. [PubMed: 1937055]
- Denzin LK, Voss EW Jr. Construction, characterization, and mutagenesis of an anti-fluorescein single chain antibody idiotype family. *J. Biol. Chem*. 1992; 267:8925–8931. [PubMed: 1577730]
- Di Camillo B, Toffolo G, Nair SK, Greenlund LJ, Cobelli C. Significance analysis of microarray transcript levels in time series experiments. *BMC Bioinform*. 2007; 8(Suppl. 1):S10.
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*. 2000; 11:4241–4257. [PubMed: 11102521]
- Gelmi, CA. <http://www.che.udel.edu/systems/people/cgelmi/tools.htm>

- Hong E, Davidson AR, Kaiser CA. A pathway for targeting soluble misfolded proteins to the yeast vacuole. *J. Cell Biol.* 1996; 135:623–633. [PubMed: 8909538]
- Huang D, Gore PR, Shusta EV. Increasing yeast secretion of heterologous proteins by regulating expression rates and post-secretory loss. *Biotechnol. Bioeng.* 2008; 101:1264–1275. [PubMed: 18781684]
- Jones EW. Vacuolar proteases and proteolytic artifacts in *Saccharomyces cerevisiae*. *Methods Enzymol.* 2002; 351:127–150. [PubMed: 12073340]
- Jonikas MC, Collins SR, Denic V, Oh E, Quan EM, Schmid V, Weibezahn J, Schwappach B, Walter P, Weissman JS, Schuldiner M. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science.* 2009; 323:1693–1697. [PubMed: 19325107]
- Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N. Statistical significance of quantitative PCR. *BMC Bioinform.* 2007; 8:131.
- Kauffman KJ, Pridgen EM, Doyle FJ III, Dhurjati PS, Robinson AS. Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 2002; 18:942–950. [PubMed: 12363344]
- Kimata Y, Ishiwata-Kimata Y, Ito T, Hirata A, Suzuki T, Oikawa D, Takeuchi M, Kohno K. Two regulatory steps of ER-stress sensor Ire1 involving its cluster formation and interaction with unfolded proteins. *J. Cell Biol.* 2007; 179:75–86. [PubMed: 17923530]
- Leek JT, Mosen E, Dabney AR, Storey JD. EDGE: Extraction and analysis of differential gene expression. *Bioinformatics.* 2006; 22:507–508. [PubMed: 16357033]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $^{-\Delta\Delta CT}$ method. *Methods.* 2001; 25:402–408. [PubMed: 11846609]
- Luo W, Chang A. An endosome-to-plasma membrane pathway involved in trafficking of a mutant plasma membrane ATPase in yeast. *Mol. Biol. Cell.* 2000; 11:579–592. [PubMed: 10679016]
- McCracken AA, Brodsky JL. Evolving questions and paradigm shifts in endoplasmic-reticulum-associated degradation (ERAD). *Bioessays.* 2003; 25:868–877. [PubMed: 12938176]
- Midelfort KS, Wittrup KD. Context-dependent mutations predominate in an engineered high-affinity single chain antibody fragment. *Protein Sci.* 2006; 15:324–334. [PubMed: 16434745]
- Moehle CM, Aynardi MW, Kolodny MR, Park FJ, Jones EW. Protease B of *Saccharomyces cerevisiae*: Isolation and regulation of the PRB1 structural gene. *Genetics.* 1987; 115:255–263. [PubMed: 3549451]
- Mori K, Sant A, Kohno K, Normington K, Gething MJ, Sambrook JF. A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. *EMBO J.* 1992; 11:2583–2593. [PubMed: 1628622]
- Mori K, Ogawa N, Kawahara T, Yanagi H, Yura T. Palindrome with spacer of one nucleotide is characteristic of the cis-acting unfolded protein response element in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 1998; 273(6):9912–9920. [PubMed: 9545334]
- Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* 2002; 20:87–90. [PubMed: 11753368]
- Nakatsukasa K, Brodsky JL. The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. *Traffic.* 2008; 9:861–870. [PubMed: 18315532]
- Newstead S, Kim H, von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.* 2007; 104:13936–13941. [PubMed: 17709746]
- Ng DT, Spear ED, Walter P. The unfolded protein response regulates multiple aspects of secretory and membrane protein biogenesis and endoplasmic reticulum quality control. *J. Cell Biol.* 2000; 150:77–88. [PubMed: 10893258]
- Nordgård O, Kvaløy JT, Farnen RK, Heikkilä R. Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: The balance between accuracy and precision. *Anal. Biochem.* 2006; 356:182–193. [PubMed: 16899212]

- Ogunnaike BA, Gelmi CA, Edwards JS. A probabilistic framework for microarray data analysis: Fundamental probability models and statistical inference. *J. Theor. Biol.* 2010; 264:211–222. [PubMed: 20170665]
- Oikawa D, Kimata Y, Kohno K. Self-association and BiP dissociation are not sufficient for activation of the ER stress sensor Ire1. *J. Cell Sci.* 2007; 120:1681–1688. [PubMed: 17452628]
- Otte S, Barlowe C. Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. *Nat. Cell Biol.* 2004; 6:1189–1194. [PubMed: 15516922]
- Parekh RN, Shaw MR, Wittrup KD. An integrating vector for tunable, high copy, stable integration into the dispersed Ty delta sites of *Saccharomyces cerevisiae*. *Biotechnol. Prog.* 1996; 12:16–21. [PubMed: 8845105]
- Patil CK, Li H, Walter P. Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. *PLoS Biol.* 2004; 2:E246. [PubMed: 15314660]
- Petracek ME, Longtine MS. PCR-based engineering of yeast genome. *Methods Enzymol.* 2002; 350:445–469. [PubMed: 12073329]
- Quackenbush J. Microarray data normalization and transformation. *Nat. Genet.* 2002; 32(Suppl):496–501. [PubMed: 12454644]
- Rakestraw A, Wittrup KD. Contrasting secretory processing of simultaneously expressed heterologous proteins in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 2006; 93:896–905. [PubMed: 16333864]
- Rakestraw JA, Sazinsky SL, Piatetsi A, Antipov E, Wittrup KD. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 2009; 103:1192–1201. [PubMed: 19459139]
- Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* 2003; 339:62–66. [PubMed: 12618301]
- Rizzo MA, Springer GH, Granada B, Piston DW. An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.* 2004; 22:445–449. [PubMed: 14990965]
- Robinson AS, Hines V, Wittrup KD. Protein disulfide isomerase over-expression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. *Biotechnology (NY)*. 1994; 12:381–384.
- Romanos MA, Clare JJ, Brown C. Culture of yeast for the production of heterologous proteins. *Curr. Protoc. Protein Sci.* 2001 May; Chapter 5, Unit5 8.
- Rothblatt J, Schekman R. A hitchhiker's guide to analysis of the secretory pathway in yeast. *Methods Cell. Biol.* 1989; 32:3–36. [PubMed: 2691854]
- Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Güldener U, Mannhaupt G, Münsterkötter M, Mewes HW. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* 2004; 32:5539–5545. [PubMed: 15486203]
- Rutledge R, Stewart D. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. *BMC Mol. Biol.* 2008a; 9:96. [PubMed: 18973660]
- Rutledge R, Stewart D. A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR. *BMC Biotechnol.* 2008b; 8:47. [PubMed: 18466619]
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, et al. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques.* 2003; 34:374–378. [PubMed: 12613259]
- Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* 2004; 22:1567–1572. [PubMed: 15558047]
- Sherman F. Getting started with yeast. *Methods Enzymol.* 2002; 350:3–41. [PubMed: 12073320]
- Shusta EV, Raines RT, Pluckthun A, Wittrup KD. Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* 1998; 16:773–777. [PubMed: 9702778]

- Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. 1989; 122:19–27. [PubMed: 2659436]
- Storey JD, Wenzhong X, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. *PNAS*. 2005; 102:12837–12842. [PubMed: 16141318]
- Takeshige K, Baba M, Tsuboi S, Noda T, Ohsumi Y. Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *J. Cell Biol.* 1992; 119:301–311. [PubMed: 1400575]
- Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I. The YEASTRACT database: A tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2006; 34:D446–D451. [PubMed: 16381908]
- Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell*. 2000; 101:249–258. [PubMed: 10847680]
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *PNAS*. 2001; 98(6):5116–5121. [PubMed: 11309499]
- Valkonen M, Penttilä M, Saloheimo M. Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 2003; 69:2065–2072. [PubMed: 12676684]
- van Anken E, Braakman I. Versatility of the endoplasmic reticulum protein folding factory. *Crit. Rev. Biochem. Mol. Biol.* 2005; 40:191–228. [PubMed: 16126486]
- Wentz AE, Shusta EV. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* 2007; 73:1189–1198. [PubMed: 17189442]
- Wittrup KD, Benig V. Optimization of amino-acid supplements for heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol. Tech.* 1994; 8:161–166.
- Xu P, Robinson AS. Decreased secretion and unfolded protein response up-regulation are correlated with intracellular retention for single-chain antibody variants produced in yeast. *Biotechnol. Bioeng.* 2009; 104:20–29. [PubMed: 19415776]
- Xu P, Raden D, Doyle FJ III, Robinson AS. Analysis of unfolded protein response during single-chain antibody expression in *Saccharomyces cerevisiae* reveals different roles for BiP and PDI in folding. *Metab. Eng.* 2005; 7:269–279. [PubMed: 15990348]
- Zhang B, Chang A, Kjeldsen TB, Arvan P. Intracellular retention of newly synthesized insulin in yeast is caused by endoproteolytic processing in the Golgi complex. *J. Cell Biol.* 2001; 153:1187–1198. [PubMed: 11402063]

Figure 14.1. Illustration of low-copy plasmids used for heterologous protein expression of scFv 4-4-20 and UPR sensor, UPRE-GFP, whereas any UPR element can be analyzed by fluorescent intensity (Robinson Lab).

Figure 14.2. Analysis of UPR and intracellular scFv levels following induction of scFv 4-4-20 expression shows UPR initiation and intracellular scFv retention starting at ~18 h. (A) In-gel fluorescence of UPRE-GFP levels in parental strain BJ5464 (top panel) compared to overexpressed BiP (HBiP; middle panel) and co-overexpressed BiP and PDI (HBiPPDI; lower panel). Comparison of each strain at 24 h postinduction of scFv expression, denoted as BJ, HBiP, and HBiPPDI, respectively, was included on each gel. (B) Western analysis using α -FLAG antibodies. Interestingly, BJ5464 maintains the highest level of intracellular expression as compared to HBiP and HBiPPDI strains. Samples of each strain at 24 h postinduction are included on each gel in order to enable a quantitative comparison. Intensities of each Western blot were normalized to the loading control, α -Act1.

Figure 14.3.

³⁵S pulse-chase analysis of scFv 4-4-20 expression and trafficking effects in *S. cerevisiae*.

Data have been modified from Figure 6 of Xu *et al.* (2005) by permission of Elsevier.

Figure 14.4.

Decision chart for analysis of microarray data. Analysis starts with differential expression testing and the method chosen for this analysis depends on the type of experiment. Differentially expressed genes should then be classified according to the Gene Ontology of the Functional Catalog, and then regulatory relationships should be investigated. For *S. cerevisiae*, the YEAstract database is a singularly good resource between transcription factors and targets or, in a more discovery-driven effort, on sequence information.

Table 14.1

Primary and secondary antibody recommendations

Identification	Vendor	Catalog number
<i>Primary antibodies</i>		
Mouse α -FLAG, monoclonal antibody	Sigma-Aldrich®	F1804
Mouse α -myc [9E10], monoclonal antibody	abcam®	ab32
Rabbit α -HA, polyclonal antibody	invitrogen™	71-5500
Rabbit α -GFP, polyclonal antibody ^a	abcam®	ab6556
Rabbit α -RFP, polyclonal antibody	abcam®	ab62341
<i>S. cerevisiae endogenous proteins</i>		
Rabbit α -Kar2/BiP, polyclonal antibody ^b	Robinson Lab	
<i>Loading controls</i>		
Mouse α -beta Actin, monoclonal antibody ^c	abcam®	ab8224
Rabbit α -GAPDH, polyclonal antibody	abcam®	ab9485
<i>Secondary antibodies</i>		
Alexa Fluor® 488 goat α -mouse IgG (H + L)	invitrogen™	A-11029
Alexa Fluor® 633 goat α -rabbit IgG (H + L)	invitrogen™	A-21070
Alexa Fluor® 546 donkey α -rabbit IgG	invitrogen™	A-10040
Alexa Fluor® 633 goat α -mouse IgG (H + L)	invitrogen™	A-21050
Alexa Fluor® 546 donkey α -goat IgG (H + L)	invitrogen™	A-11056

^a Recommend using the α -GFP rabbit polyclonal antibody ab6556. ab290 and ab6556 are more specific than ab1218; meanwhile, ab6556 is the purified version of ab290.

^b Rabbit polyclonal antibodies were generated (Capralogics) against a purified MBP-KAR2 fusion protein that was constructed from the C terminal region of KAR2 (*Xba*I site through *stop* codon) and inserted into pMAL-c2G (New England Biolabs). α -Kar2/BiP was screened for specificity in yeast lysates and purified from serum using a Montage® Antibody Purification kit with PROSEP® -A Media (Millipore) and stored at -80°C .

^c ab8224 works well as a loading control for *S. cerevisiae*. We recommend using ab8224 instead of the alternative ab8226.