# Project Summary (one page)

## 1. Overview

### (a) Title: Rational Protein Stabilization – Evaluation of Strategies and Targets

### (b) Senior personnel: Dr. Christine Vogel – PI (NYU); Dr. Dennis Shasha – Co-PI (NYU)

### (c) Activity:

**Premise.** Our goal is to manipulate protein expression levels through changes in protein stability. However, altering the stability of entire proteins *in vivo* has been drastically hindered by the lack of systematic information on protein degradation rates and their changes under different conditions, a thorough understanding of sequence and structure features impacting degradation, and our inability to clearly isolate protein ubiquitination events that lead to degradation from those with other outcomes.

**Proposed work.** We have designed a strategy to address this question by taking advantage of a highly focused and specific system that explores protein degradation in response to oxidative stress. We benefit from the use of (i) the well-characterized yeast as a model system; (ii) the combination of enrichment systems, inhibitors and large-scale proteomics, which deliver quantitative information on modification events degradation and degradation rates; and (iii) precisely timed stress treatment to maximize the observed specificity of protein modification and degradation regulation.

**Aim 1A.**  Using pulsed metabolic labeling and quantitative mass spectrometry, we will estimate the relative degradation rates for several hundreds to thousands of proteins under steady state conditions and in response to H2O2 treatment. To learn which of the changes in rates may be caused by ubiquitination, we will repeat the experiments but inhibit global ubiquitination with PYR-41. **Aim 1B.** Using metabolic labeling and ubiquitination/oxidation enrichment systems, and mass spectrometry, we will identify proteins whose specific modification patterns change in response to H2O2. Proteins that are found to be ubiquitinated (**1B**) and change degradation when ubiquitination is globally inhibited (**1A**) are strong candidates for ubiquitination-dependent protein degradation and primary targets for future manipulation. Oxidation propensity for each protein will be monitored as a protein feature that potentially influences stability and ubiquitination.

**Aim 2A.** Building on an extensive set of sequence features, a database of homology-modeled protein structures, and computational data mining, we will compile a comprehensive, genome-wide list of features associated with possible protein modification and subsequent degradation. **Aim 2B.** We will integrate the computational (**2A**) and experimental (**1A,B**) data into a model that, using a ‘Regression Trees’ algorithm, learns for each observed protein which sequence and structure features are predictive of its change in degradation rate under oxidative stress, and the ubiquitination and oxidation dependence of this change. The algorithm will extract sequence and structure features that are highly predictive across all proteins and possible modification signals. **Aim 2C.** Using the model from **aim 2B**, we will predict the degradation behavior and modification dependence for proteins not observed in the primary datasets obtained in **aim 1A,B**. The predictions will be validated by targeted mass spectrometry.

## 2. Intellectual merit (Transformative aspects)

The proposed work provides the first comprehensive and quantitative evaluation of eukaryotic protein modification and degradation under oxidative stress (aspects of which have been the subject of long-standing debate). We will use state-of-the-art molecular tools and mass spectrometry to conduct highly specific, genome-wide and highly representative experiments. We have established a unique system which allows us to trace protein modifications and degradation with high precision. The unbiased, system-wide nature of the project vouches for the likely identification of generalizable principles that can be taken to other organisms and systems. It will also allow us to identify novel targets for future manipulation of protein degradation and modification under stress.

## 3. Broader impacts

**(a**) First, the advanced proteomics, experimental, and computational techniques involved will provide inter-disciplinary training for several students and postdoctoral researchers. The PI will also prepare lectures for high-school students who are part of the American Museum of Natural History’s educational program (LANG program). **(b) Keywords -** regulatory network; predictive model; cellular behavior; protein degradation; oxidative stress; proteomics; ubiquitination; proteasome

(PROPOSAL total: 5-8 pages)

## 1. Objectives

## REFS on this page

**Overarching goal.** Our goal is to manipulate protein expression levels through changes in protein stability. However, altering the stability of entire proteins *in vivo* has been drastically hindered by the lack of systematic information on protein degradation rates and their changes under different conditions, a thorough understanding of sequence and structure features that impact degradation, as well as our inability to clearly isolate protein ubiquitination events that lead to degradation from those leading to other outcomes. We have designed a strategy to address this question by taking advantage of a highly focused and specific system that explores protein degradation in response to oxidative stress.

**Aim 1A - Rates.**  Using pulsed metabolic labeling and quantitative mass spectrometry, we will estimate the relative degradation rates for several hundreds to thousands of proteins under steady state conditions and in response to H2O2 treatment. To learn which of the changes in rates may be caused by ubiquitination, we will repeat the experiments but inhibit global ubiquitination with PYR-41.

**Aim 1B - Modification.** Using metabolic labeling and ubiquitination/oxidation enrichment systems, and mass spectrometry, we will identify proteins whose specific modification patterns change in response to H2O2. Proteins that are found to be ubiquitinated (**1B**) and change degradation when ubiquitination is globally inhibited (**1A**) are strong candidates for ubiquitination-dependent protein degradation and primary targets for future manipulation. Oxidation propensity for each protein will be monitored as a protein feature that potentially influences stability and ubiquitination.

**Aim 2A - Features.** Building on an extensive set of amino acid sequence features, a published database comprising the entire set of homology-modeled protein structures from yeast, and computational data mining, we will compile a comprehensive list of protein sequence and structure features associated with possible protein modification and subsequent degradation. The dataset will be genome-wide and can be extended as needed.

**Aim 2B - Model.** We will integrate the computational (**2A**) and experimental (**1A,B**) data into a model that, using Boosted ‘Regression Trees’ (generalized decision trees) algorithm [1], learns for each experimentally observed protein which sequence and structure features are predictive of (i) its change in degradation rate under oxidative stress, and (ii) the ubiquitination and oxidation dependence of this change. The algorithm will extract sequence and structure features that are highly predictive across all proteins, and we will screen these for possible modification signals. Our preliminary results have shown that lysines, residues known to be ubiquitinated, are among the most predictive features.

**Aim 2C - Validation.** Using the model from **aim 2B**, we will predict the degradation behavior for proteins not observed in the primary datasets obtained in **aim 1A,B**. The model will predict if these proteins are degraded under oxidative stress, and if this degradation is likely dependent on protein modifications. The prediction of protein degradation will be validated by targeted mass spectrometry experiments. Targeted mass spectrometry allows for higher sensitivity in protein and peptide detection, as well as more accurate quantification – but the approach is limited to tens to few hundreds of proteins, due to the need of *a priori* knowledge of peptides to be measured.

### Contributions and transformative aspects (particularly suitable for the EAGER mechanism)

The tight regulation of protein degradation via different degradation pathways under oxidative stress render this system an ideal starting point to address fundamental biological questions: **how is protein degradation regulated in response to a stimulus and, which protein sequence and structure features determine the fate and pathway that an individual protein will undergo?**

**=> Solving a debate.** The different routes of protein degradation upon stress, i.e. ubiquitin-dependent or –independent pathways (**Fig. 1.1**), are a matter of ongoing debate. The use of recently developed molecular tools and state-of-the-art mass spectrometry techniques will allow us reveal the relative contributions of different degradation pathways to the oxidative stress response and also highlight regulatory signatures that determine them.

**=> Understanding a dynamic system.** Only few large-scale datasets on protein stability exist to date [2-5], and we have little understanding of how protein stability changes (in a protein-specific manner) in response to stimuli. We will examine a *dynamic* system, i.e. the oxidative stress response, and accurately *quantify* proteins and peptides using high-resolution mass spectrometry and provide *large-scale* datasets of protein degradation, oxidation, and ubiquitination.

**=> Constructing an unbiased, predictive model.** We will use *scalable modeling* techniques that can handle both sequence and structure features in a linear and non-linear way. The unbiased, system-wide nature of the project vouches for the identification of generalizable principles that can be taken to other organisms and systems. It will also allow us to identify novel targets for future manipulation of protein degradation and modification under stress.

|  |
| --- |
|  |
| **Fig. 1.1. Different paths of protein degradation under oxidative stress** |

**=> Predict testable modification strategies.** We will directly *test* the predictive model via targeted assays on selected proteins.

**An example of success.** An example of successful regulation of protein degradation is the transcription factor Yap1 which regulates several stress related genes [6]. A Yap1 mutant, in which the cysteine-rich N-terminal domain (containing the E3 ligase interaction site) was deleted, rendered the protein more stable and the cells less sensitive to stress [6] – directly connecting protein degradation regulation and an ‘improved’ stress response. Our preliminary model already successfully predicts the degradation propensity of Yap1 confirmed by this study (**Prelim. Results**).

### Integration into a future research program

The proposed work provides key building blocks for the next steps in understanding and modifying protein degradation. (i) With highly sensitive, but experimentally non-trivial site-specificubiquitination assays, we can extract *local* information on likely ubiquitination sites and their sequence contexts. Given that the average protein contains >25 lysines (all of which can be ubiquitinated), the method will refine the predictions based on the data from **aim 1B** above. (ii) Protein oxidationunder stress provokes structural changes and can trigger protein degradation. A set of future *site*-specific experiments can quantify local oxidation evens and provide data to include into the predictive model (**2B**). (iii) Once modification sites (in particular lysines predicted to be ubiquitinated) have been identified and validated, the next step is to genetically manipulate degradation of the respective protein through site-specific mutagenesis. Such experiments, when conducted on stress-defense proteins, can then be evaluated in their protective effect on the cell, i.e. through increases in viability. Manipulating gene expression levels through protein stabilization or destabilization provides new strategies for bioengineering and plants sciences. For example, during crop storage, seeds are exposed to environmental stresses which affect germination capability and yield. Stabilizing key proteins of the stress response can render cells less sensitive to these stressors and increase the overall cellular half-life.

## 2. Background

**Protein degradation as a crucial part of gene expression regulation and follows different pathways.** Proteasomal degradation accounts for >90% of cellular protein turnover [7], and failure to degrade oxidatively damaged proteins has detrimental effects for any cell [8]. Despite decades of intense study, the exact role of ubiquitination for removal of these damaged proteins is the subject of an ongoing debate. Poly-ubiquitination through lysine 48 (K48) is the most common modification leading to protein degradation [9]. It allows the cell to target proteins for degradation in a highly specific manner [10], but requires strict regulation and cellular energy. While ubiquitin-dependent degradation of oxidized proteins has been observed [11-14], many proteins are also removed via ubiquitin-independent pathways [15-18] (**Fig. 1.1**). Ubiquitin-independent degradation does not require tagging, and little is known about the recognition of untagged proteins for degradation. The specific *in vivo* protein substrates of each pathway, ubiquitin-dependent or –independent, have not been characterized.

**Ubiquitination is complex with many different roles.** Poly-ubiquitination is a frequent and evolutionarily conserved [19-21] protein modification that (through different lysine linkages) can have a variety of cellular functions. Consequently, it has been very difficult to determine exact ubiquitination motifs and, even less so, their usage under different conditions, e.g. for protein degradation.Several studies have identified global ubiquitination sites [22, 23], and it is thought that the regions around the ubiquitinated lysine are often depleted of arginine on the N-terminal side and lysine and histidine on both sides. Ubiquitinated lysines may prefer sequence neighborhoods with negative net charge[22-24]. In addition, poly-ubiquitination sites can be impacted by the protein’s N-terminus [25] and tend to occur towards structured regions [23]. Large-scale ubiquitination studies have so far only examined steady-state conditions, identifying global sequence signatures of ubiquitination [22, 23]. Despite the use of inhibitors and other approaches [22], it is not clear which of these ubiquitination events trigger degradation and which have other functions, nor is it clear under which conditions these roles may change.

|  |
| --- |
|  |
| **Fig. 2.1**. **Objective:** Building and testing a predictive model of protein degradation under oxidative stress*.* |

**Manipulating protein degradation is difficult.** Since protein degradation is an important part of gene expression regulation, researchers have attempted to manipulate gene expression levels via this route. However, these attempts have proven to be challenging. Some studies have attempted to predict protein stability (or degradation) from sequence and structure, but little is known about changes to protein stability in response to stimuli, e.g. stress. Some amino acids are thought to destabilize a protein if at the N-terminal end [25, 26]; also proline, glutamate, serine, and threonine stretches can cause degradation[27]. Proteins with many intrinsically unstructured regions also degraded more rapidly than other proteins [28]. The proteasome itself has chaperone activity and can recognize un-/misfolded proteins [29, 30] – but how exactly these sequence and structure features are recognized is unknown. Changing the N-terminal residue or sequence stretch has been successfully used to stabilize a protein [31, 32].

Likewise, modifying a protein’s ubiquitination state and hence its putative ubiquitin-dependent degradation has been very challenging. Typical experiments involve the step-wise substitution of all lysine residues by arginine [33, 34]. Due to the lack of a clear sequence signal, the experiments have operated on a trial-and-error basis with mixed success.

**Protein degradation is highly regulated under oxidative stress.** Oxidative stress is characterized by an imbalance between reactive oxygen species (ROS) and cellular antioxidant defense. Some proteins appear to be more sensitive to oxidative stress than others, but the specific oxidative propensity for each protein, as a function of its sequence and structure, is unknown. The oxidation of amino acids, primarily proline, arginine, lysine, threonine, glutamate, and aspartate generates carbonyl groups [35] that are extremely common [36, 37].

Despite the existence of stress defense and damage repair mechanisms [38, 39], the majority of oxidatively damaged proteins have to be removed from the cell to re-establish cellular proteostasis. Accumulation of oxidized proteins and toxic aggregates causes cellular senescence, loss of replicative ability and cellular death in all organisms [40], for example, impacting plant seed germination [41, 42] and causing diseases [43, 44]. The highly efficient, regulated, and dynamically adaptive degradation of proteins is an essential part of the oxidative stress response. Contradictory evidence exists on the requirement for ubiquitination of oxidized proteins (**Fig. 1.1**). Several studies report that the 20S proteasome (and the mammalian immunoproteasome) degrades structurally abnormal, misfolded, or highly oxidized proteins in a ubiquitin-independent fashion [16, 18, 45-49], while other studies highlight the importance of ubiquitination during the oxidative stress response [12-14, 50].

**Our strategy.** We have designed a strategy to address this question by taking advantage of a highly focused and specific system that explores protein degradation in response to oxidative stress (**Fig. 2.1**). The analyses will produce the large-scale datasets needed to further our understanding of protein degradation regulation, but also provide the first computational model that quantitatively predicts protein degradation and correlative protein features. Using this system, we will identify, for thousands of proteins, if degradation rates change in response to oxidative stress and if the proteins are ubiquitinated or oxidized. Based on this data and a comprehensive collection of sequence and structure based protein features, we will computationally learn to predict a protein’s fate with respect to degradation. We will use and validate the model on a set of proteins which have not been characterized in the experiments, but whose functions suggest substantial roles in the oxidative stress response.

## 3. Investigator qualifications and preliminary data

|  |
| --- |
|  |
| **Fig. 3.1. Ubiquitin-K48 accumulates under oxidative stress together with oxidized proteins in the presence of proteasome inhibition** (red box) suggesting a role for ubiquitin-dependent degradation. |

Our team comprises a unique combination of expertise in proteomics, molecular biology, computational data analysis, as well as the use of yeast as a model system for the oxidative stress response. Dr. Christine Vogel (PI) has extensive expertise in quantitative proteomics [51-54], protein sequence and structure analysis [55-59], and computational data analysis [60-63]. Dr. Dennis Shasha (CoPI) is an expert in biological data mining using mathematical approaches [64-71]. Dr. Gustavo Silva(postdoctoral researcher) has published extensively on proteasome activity under oxidative stress and the ubiquitin/oxidation proteomic system [34, 72, 73].

### Published work

**Quantifying proteins at large scale (*Nature Biotech 2007,25(1):117*; *Nature Protocols 2008*)*.*** We developed a mass spectrometry based method, called APEX (Absolute Protein EXpression index) that estimates absolute concentrations for thousand of proteins in complex samples and identifies statistically significant differential protein expression [51, 52].

**Protein translation and degradation substantially affect gene expression (*Nat. Rev. Genetics 2012, 13(4):227; Mol. Sys. Bio. 2010,6:400; Proteomics 2010,10(23):4209*).** We demonstrated that protein concentrations are highly conserved across species – and, on average, the protein concentrations of orthologs correlate better between two species, than the corresponding mRNA concentrations [74]. Further, using computational modeling on quantitative mass spectrometry data of >1,000 human proteins and ~150 sequence/structure features, we showed that transcription can only explain a small fraction (27%) of the variation in concentrations of human proteins, and translation and protein degradation account for a larger portion of the variation (~40%).

**Protein degradation and translation are heavily regulated during oxidative stress (*Mol & Cell Proteomics, 2011,10(12)*)*.*** Using APEX-based quantitative mass spectrometry, we analyzed time series data from >1,700 yeast proteins [51, 52, 54] and demonstrated that proteins have very different dynamics from RNA during the oxidative stress response indicating extensive regulation at the protein level: while RNA reacts strongly within the first 30min and then returns to normal levels, protein concentrations change at a slower rate but maintain these changes in the time scale of the experiment.

### Pilot studies

**Both ubiquitin-dependent and –independent protein degradation occurs under oxidative stress.** We have successfully established the experimental model used in this proposal. In permeabilized yeast, we inhibit ubiquitination and proteasomal degradation by use of PYR-41 and MG-132, respectively (**Fig. 3.1**). While some have argued that ubiquitination is not required for degradation of oxidized proteins [15, 16, 18], our results suggest that a specific proteins are regulated through K48 ubiquitin-dependent degradation. Both oxidized and ubiquitinated proteins accumulate immediately after H2O2 treatment in yeast. The majority of oxidized proteins are not removed in the presence of the proteasomal inhibitor MG-132, corroborating that their degradation is dependent on proteasome activity (**Fig. 3.1**)[8].

However, oxidation and ubiquitination display very different dynamics during later stages of the experiment (*recovery*): while both oxidation and ubiquitination immediately increase after *stress* induction, ubiquitinated proteins disappear faster than oxidized proteins (**Fig. 3.1**). K48-linked ubiquitination (leading to degradation) strongly accumulates after ~2hrs. Polyubiquitin linked through K48 accumulates concomitantly to the disappearance of oxidized proteins.

Moreover, oxidized proteins seem to be largely tagged by K48-linked polyubiquitin, indicating their ubiquitin-dependent degradation (**Fig. 3.2**). Similarly, when ubiquitination is globally inhibited by PYR-41, degradation of oxidized proteins is dramatically slowed down. These results suggest that (i) both ubiquitin-dependent and –independent degradation occurs in response to oxidative stress, and (ii) ubiquitin-dependent degradation dominates the response at about 2 to 4hrs *recovery*. We carefully designed the proposed experiments based on these results to isolate the K48 linked ubiquitin activity.

**Mass spectrometry identified oxidized, ubiquitinated proteins, and global protein expression.** We employed label-free APEX-based mass spectrometry to estimate abundances of proteins in the whole cell extract and to identify proteins that are oxidized, or ubiquitinated (*not shown*). We identified ~2,100 proteins in their H2O2 *stress* and 2hr *recovery* response compared to control. While this data is only semi-quantitiative, we already find that different proteins follow very different pathways of ubiquitination, oxidation, and degradation. For example, ribosomal proteins are highly oxidized and ubiquitinated, leading to a decrease in protein levels during the recovery. In contrast, stress and folding related proteins are also highly ubiquitinated and oxidized, but increase in expression.

**Stochastic gradient descent based modeling can identify predictive sequence features.** Despite the semi-quantitative nature of the datasets above and the lack of site-specific information, a preliminary model using the stochastic gradient descent algorithm (**aim 2**) identified a set of *global* features predictive of degradation, ubiquitination, and oxidation (**Tab. 3.1**). Interestingly, even in this first model, the lysine content of proteins is amongst the predictive features.

The preliminary model already highlights interesting examples and putative targets for modification. Yap1, the stress-related transcription factor mentioned above, is not in the *observed* experiment, but based on its sequence the model predicted it to degrade under stress, consistent with published work [6]. Thus, even the preliminary model can recapitulate biological observations. Furthermore, sulfiredoxin (Srx1), a key enzyme of the stress defense, was predicted as a strong candidate for ubiquitination/degradation by the model (**Tab. 3.1**). Srx1p is in fact known to be degraded under oxidative stress [39]. The most N-terminal lysine of Srx1 is negatively charged and may be the ubiquitination site (*not shown*). Since the Srx1 deletion mutant is stress sensitive [39], we hypothesize that stabilization of the protein through removal/mutation of the ubiquitination site may specifically increase stress tolerance. **Aim 3** of this proposal addresses testing predictions of the model, including examples like Yap1 and Srx1.

**Tab. 3.1. Preliminary model.** Most predictive sequence features, listed in their informal description

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Positive** | **Negative** | **Top three protein predictions** |
| **Degradation** | Lys, Pro, PEST motif | Codon adaptation, Arg, Isoelec point | Shm1, Rli1, Hhf1 |
| **Ubiquitination** | Codon adaptation | Disorder , Lys, Length | Idh1, Srx1, Nop16 |
| **Oxidation** | Codon adaptation, Ile | Ser, Disorder, Length | Sec53, Cpa1, Por1 |

## 4. Research methods

**Abbreviations. DEG** – degradation; **MS** – mass spectrometry; **OX** – oxidation; **UB** – ubiquitination

### Biological system and supporting infrastructure

**Strains.** All experiments will be conducted with the *S. cerevisiae* strain RJD1171 (MATa his3Δ200 leu2-3,112 lys2-801 trp1Δ63 ura3-52RPT1FH::Ylplac211 (URA3)[75]) that has one proteasome component (Rpt1) tagged with the FLAG epitope and a poly-histidine tail. Since the strain is not auxotrophic for lysine and arginine (which is a requirement for SILAC), we will create a K/R-auxotroph RJD1171 strain or use N15 labeling for quantitative mass spectrometry. Alternatively, we will use the MBS164 strain (MATa, ura3-52, leu2Δ1, trp1-63, his3-200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, arg4Δ, sml1::TRP bar1::HIS3) derived from S288C [76] which is a K/R-auxotroph.

**Growth conditions.** To enable efficient inhibitor uptake (see below), the strains are grown under conditions that induce permeability of the plasma membrane, i.e. in Minimal Proline Dextrose (MPD) medium [77]. Cells are allowed to divide at least 6 times and are treated in log phase (OD600~0.2). The cells are incubated with 0.003% SDS for 90 min prior to the inhibitors treatment for 15 min [78](**Fig. 4.1A**). The SDS treatment does not interfere with cell growth [77]. Our preliminary studies have shown that accumulation of oxidized proteins is achieved by a 45 min pulse treatment with 0.6 mM H2O2; cells are then transferred to fresh medium also containing SDS (and inhibitors if specified) for recovery.

**Stress treatment.** Our **Prelim. results** demonstrate that both UB-dependent and –independent protein degradation occurs in very specific protein groups, and maximal specificity for degradation linked ubiquitination (K48) is reached at four hours recovery. We will conduct all stress experiments at this time point.

**Proteomics.** For all mass spectrometry experiments, we will use an LTQ Orbitrap Velos (Thermo) coupled to a 2D nano-flow liquid chromatography system (Eksigent) available in the Vogel lab. This setup is ideal for the proposed experiments, as the high sensitivity combined with automated high speed sample analysis allows for efficient quantitation of complex protein samples.

|  |
| --- |
| NEW – Gustavo? |
| **Fig. 4.1. Overview - Experiments. (A)** Experimental setup, basic time course experiment, conditions, and main samples (*control, stress, recovery*). **(B)** Measuring degradation at *recovery* with pulsed-SILAC, small time course proteomics experiments. **(C)** Measuring oxidation and ubiquitination at *stress* and *recovery* by standard SILAC.  |

**Sample preparation and analysis.** Protein samples will be prepared as described before [54]. Briefly, cells are lysed, and lysate will be reduced, cysteines alkylated with iodoacetic acid, digested with trypsin and cleaned by C18 filtering. To maximize proteome coverage, samples will be fractionated using the OFFGEL fractionator (expedeon) CV: NEEDED? and subjected to reverse phase liquid chromatography using a three-hour 5-60% acetonitrile gradient on a Agilent Zorbax C18 column. Three biological replicates will each be analyzed twice (technical replicates) in identical mass spectrometry runs [60]. All MS/MS data will be mapped to protein sequences and quantified using established pipelines, including those established in our lab [51, 52, 79-84].

**Stable Isotopic Labeling of Amino acids in Cell culture (SILAC) to quantify proteins and peptides.** Since the respective UB and OX enrichment/immunoprecipitation steps (see below) modify sample constitution and introduce errors, the use of label-free techniques only provides semi-quantitative data. The solution we offer is the use of isotopically labeled amino acids where protein concentrations are quantified through comparison of intensities of ‘heavy’ and ‘light’ versions of a peptide spectrum between samples [85]. The SILAC technique is very well established with standard protocols [85, 86]. In brief, cells will be grown in medium containing isotopically labeled L-lysine and L-arginine for several generations to maximize label penetrance. The choice of lysine and arginine relies on the fact that each tryptic peptide will contain at least one modified amino acid, enhancing coverage of quantifiable peptides. Heavy and light amino acids will be used for experiment and control, respectively. Alternatively, we may use N15 labeling of the nitrogen atoms in the backbone of the peptide chain.

**Table 4.1. Time line**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **1** |  |  |  | **2** |  |  |  |
| Aim 1A – Protein degradation under normal and stress conditions | x | x | x | x | x |  |  |  |
| Aim 1B – Protein ubiquitination and oxidation under normal and stress conditions  |  | x | x | x | x | x | x |  |
| Aim 2A – A compendium of genome-wide protein sequence and structure features | x | x | x | x | x |  |  |  |
| Aim 2B – A predictive model of protein degradation under oxidative stress |  |  | x | x | x | x | x |  |
| Aim 2C – Targeted validation of predicted protein degradation changes under stress |  |  |  |  | x | x | x | x |

### Specific aims

### Aim 1A – Protein degradation (and its dependence on ubiquitination)

**Goal and expected outcomes.** Despite the important role during protein expression regulation [5, 60], estimates of protein degradation rates are still rare and often exist only for cells growing under normal conditions [2]. Using pulsed metabolic labeling and quantitative mass spectrometry, we will estimate the relative degradation rates for several hundreds to thousands of proteins under steady state conditions and in response to H2O2 treatment. Further, we will investigate if protein-specific degradation is dependent on ubiquitination or not. The analysis will provide a quantitative answer to the long-debated question if ubiquitination is required for degradation of oxidized proteins.

**Estimating protein degradation with pulsed-SILAC (Fig. 4.1B).** We will derive and compare estimates of protein degradation *rates* (or the opposite: protein stability) for two conditions: H2O2 treated cells and the equivalent time point in an *untreated* control (**Fig. 4.1B**). To learn which of the changes in rates may be caused by ubiquitination, we will repeat the experiments but inhibit global ubiquitination with PYR-41. A labor- and material-intensive, but also most accurate and least invasive method is pulsed-SILAC [87], a method which follows, in a time course experiment, disappearance of labeled proteins after exposure to a pulse of isotopically labeled amino acids. Protein concentration changes will be estimated from ion intensities [5] using standard software [79, 81]. The degradation rate represents the slope of a curve fitted through the data [87] (**Fig. 4.1B**). If DEG=*rate\_H2O2/rate\_untreated* >0, the protein degrades faster under stress, and vice versa.

**Validation, challenges, extensions.** Alternative methods are simpler but more invasive. We can inhibit translation with cycloheximide, such that a decrease in protein concentrations (estimated by label-free APEX-based proteomics [51, 88]) is purely due to degradation. However, inhibiting translation perturbs the cells substantially. To monitor the extent of proteasome inhibition and alternative degradation pathways, we will also quantify proteasome activity using an *in vitro* fluorescence assay [72]. Western blots and targeted MS experiments will serve to validate expression changes of individual proteins.

### Aim 1B – Protein ubiquitination and oxidation under normal and stress conditions

**Goal and expected outcomes.** Using SILAC, ubiquitination/oxidation enrichment systems, and mass spectrometry, we will identify proteins whose specific modification patterns change in response to H2O2. We will quantify protein ubiquitination and oxidation under normal and stress conditions. Proteins that are found to be ubiquitinated (**1B**) and change degradation when ubiquitination is globally inhibited (**1A**) are strong candidates for ubiquitination-dependent protein degradation and primary targets for future manipulation. Oxidation propensity for each protein will be monitored as a protein feature that potentially influences stability and ubiquitination.

|  |
| --- |
| UPDATE FIGURE (Gustavo?) – ONE FIGURE 4.1 and 4.2?! |
| **Fig. 4.2. Quantifying ubiquitination and oxidation at the peptide level**. Samples will be enriched and quantified by standard SILAC.  |

**Quantifying ubiquitination and oxidation (Fig. 4.1C, 4.2).** To quantify ubiquitination and oxidation, samples will be combined from SILAC experiment (stress vs. control). Ubiquitinated and oxidized proteins will be enriched and separated from unmodified proteins via protocols we already established in the lab. Ubiquitinated proteins will be precipitated via the TUBE system (*COMPANY*) which comprises arrays of Tandem Ubiquitin Binding Entities. To isolate oxidized proteins, we will exploit one of the most common forms of protein oxidation affecting several amino acids [89] – carbonylation. Carbonylated proteins will be derivatized with 2,4-dinitrophenylhydrazine (DNPH), and immunoprecipitated using magnetic Dynabeads Protein G (Invitrogen) loaded with anti-DNP antibody (Sigma)(**Fig. 4.2**). After enrichment, we will use LC-MS/MS to identify and quantify modified peptides and proteins. Changes in ubiquitination or oxidation will be monitored through differential intensities of the corresponding peptides. This approach is well-established [22, 23] and will provide protein-specific information on the modification.

**Validation, challenges, extensions**. As shown in the **Prelim. results**, the time points chosen here primarily display K48 ubiquitination leading to protein degradation. Protein ubiquitination can be validated in gel digestion/MS of well-defined molecular weight ranges and through literature and databases [90-93](e.g. http://scud.kaist.ac.kr/). **–** Biotin-tagged DNPH is available as an alternative for the anti-DNP antibody. We will use western blots and literature [94, 95] to verify protein oxidation.

### Aim 2A – A compendium of genome-wide protein sequence and structure features

**Goal and expected outcomes**. Sequence and structure determine the fate of a protein. Exposed residues may be preferentially oxidized, intrinsically disordered proteins may be more rapidly degraded than other proteins [28]. While a few ubiquitination signals have been identified [22, 23], their role in triggering protein degradation upon oxidative stress is still unknown. Building on an extensive set of amino acid sequence features, a published database comprising the entire set of homology-modeled protein structures from yeast, and computational data mining, we will compile a comprehensive list of protein sequence and structure features associated with possible protein modification and subsequent degradation. The dataset will be genome-wide and can be extended as needed. A mapping of modifications (e.g. ubiquitination) to sequence and structure features will provide first insights into possible motifs and amino acid enrichments that provide regulatory signals. In assembling these features, we will exploit our extensive expertise in protein sequence and structure analysis [55-60, 96-102].

SHORTEN – features are discrete!!!

**Sequence features** will include: sequence length, percentage of secondary structures (which impacts protein folding and stability), percentage of hydrophilic and hydrophobic, buried and surface amino acid residues (describing protein folding and residue exposure at the surface), the presence of PEST-rich regions and specific N-terminal residues (which are degradation signals [26, 103]), percentage of oxidation-prone residues at the protein surface, vicinity of putatively oxidized residues to lysines (which may be ubiquitinated), or intrinsic unstructuredness (which has been shown to strongly anti-correlate with protein stability [28, 60]). Amino acids (and sequence stretches) will also receive a conservation score, i.e. a score describing how much the residue (or sequence stretch) is conserved across evolution, exploiting the fact that functionally relevant sequences are likely conserved across organisms.

**Structural features** have been shown to substantially help prediction of post-translational modification sites [104]. UPDATE CHRISTINE – Rich’s database The features used here will include:

1. Using DisEmbl and DisoPred[105, 106], we will predict the propensity of sequence stretches to assume specific secondary structures, such as alpha helices, beta strands, or coils.
2. Using protein domain structure predictions, e.g. InterPro or SUPERFAMILLY [107, 108], we will determine which amino acid residues located in which secondary structure elements and also if these residues are at the surface or buried in the core of the protein domain structure.
3. Using existing protein three-dimensional structures deposited in PDB and SCOP [109-111], we will extract exact data on the position of amino acid residues within secondary structure elements and within the protein domain structure. These data can serve as a high-confidence dataset to validate the predicted data from b).

**Modification sites**. Since we know that ubiquitination and oxidation occur at specific amino acid residues (K and mainly P, R, K, T, E, D, respectively), we will derive a set of features that examine sequence/structure properties adjacent to these modified residues in a window of +/- ten amino acids. Some of these features will be similar to the ones above describing global sequence and structure, but others will be position-specific. Further features will describe the relationship of the 20 amino acid stretch to the entire protein, e.g. its position within the protein structure (exposed/buried) or sequence (towards N- or C-terminal end). To include these site-specific features in the global protein analysis (above), we will construct composite features describing the *average* number of ubiquitination sites per protein and the *average* site-specific properties.

### Aim 2B – A predictive model of protein degradation under oxidative stress

DENNIS – CORRECT...

**Goal and expected outcomes.** The central goal of this proposal is to predict protein degradation. We will integrate the computational (**2A**) and experimental (**1A,B**) data into a model that, in addition to stochastic gradient descent, uses Boosted ‘Regression Trees’ (generalized decision trees). Both algorithms will learn, for each experimentally observed protein which sequence and structure features are predictive of (i) its change in degradation rate under oxidative stress, and (ii) the ubiquitination and oxidation dependence of this change. Like decision trees, regression trees [1] learn which discrete features of the input data set (in this case, sequence and structure features), predict the output (in this case, degradation rates). The algorithms are complementary in that regression trees are easier to interpret, but stochastic gradient descent handles feature interaction better. The sources of data include information about the differential importance of the many sequence/structure features, their positive or negative influence on degradation (DEG), ubiquitination (UB), and oxidation (OX). The output of the model, for a given sequence, will be an estimate of its fate by predicting the values for OX, UB, DEG. The combination of positive and negative values for OX, UB, DEG defines the protein’s pathway. For a given protein, we will identify the sequence features within the protein that strongly match those features associated with an outcome. For example, if a protein is predicted to be ubiquitinated (since its peptides are predicted to be ubiquitinated), and we observe large coefficients C\_j for a motif with hydrophobic residues, then we will examine the protein sequence for the respective hydrophobic region. This work will be carried out with Dr. Dennis Shasha (Co-PI) and build on our expertise in computational data analysis and modeling [51, 54, 56, 59-63]. The results of the modeling will, for the first time, outline general principles that govern protein degradation and specific regulatory sites and motifs for the entire yeast genome (**Fig. 4.3**).

**Basic setup.** We will model degradation as a function of discrete protein sequence and structure features assembled in **aim 1** (**Equ. 1**), using the experimental data (DEG) and protein-specific features:

[condition\_A, protein\_1, feature 1\_1, …, feature\_n\_1, DEG\_1] (1)

[condition\_A, protein\_2, feature 1\_2, …, feature\_n\_2, DEG\_2]

...

[condition\_A, protein\_m, feature 1\_m, …, feature\_n\_m, DEG\_n]

where the number of features n is on the order of 300, and the number of proteins m>3,000. Each row will have data fields in which the features and degradation rate are normalized values. We will model each protein’s ubiquitination and oxidation by substituting UB\_i and OX\_i for DEG\_i for protein\_i, respectively. UB and OX will be modeled for n entire proteins, and the feature list will include site- and position-specific features in addition to global protein features (**Equ.s 2BC**).

**Modeling.** The central modeling problem consists of the use of different algorithms to find, for each condition over all proteins, a single set of coefficients C\_i to each feature i so we can obtain equations of the form

DEG for protein\_i at condition\_A

= C\_DEG\_1 \* feature\_1\_1 + … + C\_DEG\_n \* feature\_n\_1 + constant\_DEG\_for\_condition\_A (2A)

OX for site\_j\_from\_protein\_i at condition\_A

= C\_OX\_1 \* feature\_1\_1 + … + C\_OX\_n \* feature\_n\_1 + constant\_OX\_for\_condition\_A (2B)

UB for site\_i\_from\_protein\_j at condition\_A

= C\_UB\_1 \* feature\_1\_1 + … + C\_UB\_n \* feature\_n\_1 + constant\_UB\_for\_condition\_A (2C)

where DEG, OX, UB denote degradation, oxidation, and ubiquitination, respectively. Given these equations per protein/peptide and condition, we can find the coefficients C that have large positive or negative coefficients and predict the dependent variable OX, UB, or DEG.

**Algorithm.** Boosted regression trees will be method of choice {Hastie, 2001 #1544}. ... *a few words on scalability, overfitting, parsimony, non-linear relationships, data heterogeneity and sparsity. [I don’t think we need to get super-technical here,but have dealt with this above I think. We don’t have room.]*

**Combined effects.** In addition to examining features with extremely large positive or negative values, we will also analyze (using both stochastic gradient descent and boosted regressio trees) the product terms of two feature values to possible *combined* effects and interdependencies between two features. For example, we may evaluate the *interaction* between protein length andthe degree of unstructuredness and the relationship between the two features*.* Inclusion of product terms will give us a slightly more complex formula of the form:

DEG for protein i

= C\_1 \* feature\_1 + … + C\_n \* feature n

+ D\_i\_j \* feature\_i \* feature\_j + … + D\_k\_m \* feature\_k \* feature\_m + constant (3)

Where the coefficients D describe the impact of the combined features. Initially we will combine features that have large positive or negative coefficients on their own. Some features may have weak effects on their own but strong effects in combination with other features.

**Validation and performance assessment.** Model predictions will be validated by ten-fold cross-validation, i.e. by training model parameters based on randomly selected 90% of the data and testing the predictions on the remaining 10%. We will derive estimates of prediction confidence and error, e.g. the relative error as (predicted-observed)/observed. Model predictions for individual proteins will also be validated by western blotting and by the experiments described in **aim 2C**. - Bayesian networks provide a robust and appropriate framework that can also automatically account for the stochasticity inherent to any biological measurements. However, Bayesian network modeling requires large quantities of training data relative to the number of coefficients that need to be tested.

### Aim 2C – Targeted validation of predicted protein degradation changes under stress

**Goal and expected results.** Using the model from **aim 2B**, we will predict the degradation behavior for proteins not observed in the primary datasets obtained in **aim 1A,B**. The model will predict if these proteins are degraded under oxidative stress, and if this degradation is likely dependent on protein modifications. The prediction of protein degradation will be validated by targeted mass spectrometry experiments. Targeted mass spectrometry allows for higher sensitivity in protein and peptide detection, as well as more accurate quantification – but the approach is limited to tens to few hundreds of proteins, due to the need of *a priori* knowledge of peptides to be measured.

**Protein selection.** The validation experiments will focus on stress-related proteins, a subset of which are listed in **Tab. 4.2**. While we expect to obtain data for a large fraction of the proteins listed in **Tab. 4.2**, these data may be incomplete (i.e. for a given protein, degradation may be measured, but not ubiquitination), not that reliable (if only one replicate reports the protein), or not observed at all (e.g. transcription factors of low abundance). However, using **aim 2B**’s model, we can *predict* the fate of any protein given a change in its sequence and structure. We will choose a set of 20 to 40 proteins from **Tab. 4.2.** which i) have key roles within the stress protective pathway; ii) display strong preference for one degradation pathway, in particular UB-dependent degradation – as predicted by the model (**aim 2**); iii) present a strong sequence signature (e.g. for ubiquitination), and iv) are measurable by MS as evidenced by MS data or APEX-predictions [51, 52]. We will focus on proteins degraded by the UB-dependent pathway (about half of the test set; the other parts will be proteins that are ubiquitinated but not degraded, as well as degraded via UB-independent pathways. In other words, the validation experiments will include positives and negatives for UB-dependent degradation.

**Tab. 4.2**. **Key yeast proteins with beneficial roles during oxidative stress** [39, 114-136]

|  |
| --- |
| **Class of proteins with examples** |
| **Stress related transcription factors**: YAP1, SKN7, MSN2, MSN4, RPN4, YAP2, GCN4, HAP1 MAC1  |
| **Thioredoxins:** peroxiredox.; thioredox.; thioredoxin reduct. - TSA1-2, AHP1, PRX1, DOT5; TRX1-3; TRR1-2 |
| **Glutathione system:** glutathione peroxidases; glutaredoxins; glutathione transferases; glutathione reductase - GPX-2, HYR1; GRX1-8; GTT1, GTT2 GLR1  |
| **Other enzymes:** superoxide dismutases; catalases - SOD1-2; CTT1, CTA1 |
| **Chaperones** - SSA1, CCS1, SSE1-2, UMP1, HSP12, 26, 48, 70, 78, 104, 150; SSA4, UMP1, POC1-4 |
| **Protein folding:** protein disulfide isomerase - ERO1, PDI |
| **Damage repair:** Metsulfoxide reductase; sulfiredoxin; endonuclease - MXR1, MXR2; SRX1; APN1 |
| **Ubiquitin:** ubiquitin; E2 conjugases; E3 ligases; deubiquitinase - UBI4; UBC4, UBC5, UBC6, UBC8;HUL4, UBP15 |
| **Proteasome:** 20S proteasome subunits; 26S proteasome subunits - PRE2, PRE3, PUP1…; RPT1-6, RPN1-14 |
| **Other proteases:** autophagy-related; vacuolar proteases - ATG1,7; YPS6, APE1, PEP4, PRB1 |
| **Small molecules:** glutathione synthetases; metallothioneins - GSH1-2; CUP1, CRS5  |
| **Reductive metabolism:** pentose phosphate enzy.; trehalose synth. - ZWF1, TKL1, RPE1; TPS1-3,TSL1 |
| **Cell cycle regulation** - RAD9, CDC28 |

**Targeted mass spectrometry** offers higher sensitivity and quantitation accuracy than data-dependent approaches from **aim 1**, but has lower throughput. Using targeted MS, we will validate model predictions on ubiquitination, oxidation, and degradation under oxidative stress for peptides or proteins that have not already been observed in **aim 1**. Even if, in individual cases, the peptide has been observed already in **aim 1**, targeted MS can be used to confirm the measurement. If candidate peptides have not been observed in the MS experiments in **aim 1A** due to their low abundance and are also absent from pertinent databases [137, 138], we will prioritize their analysis based on their predicted observability which we obtain from the computational sequence-based prediction that we developed [51, 52].

We will employ two different targeted MS methods which analyze individual *peptides* derived from the proteins of interest. (A) Using an LTQ Orbitrap Velos instrument (Vogel), we will analyze an *inclusion* *list* containing the masses (i.e. mass-to-charge ratios) of desired peptides. The instrument will isolate, fragment (identify), and quantify these peptides (and their modifications) in the full scan mass spectrum. In contrast to standard data-dependent methods which ignore low abundance peptides, any peptide from the inclusion list is monitored, an approach that increases the overall sensitivity the of the analysis. The method can measure hundreds of peptides in one analysis and can use the SILAC samples from **aim 1A**. (B) For a smaller subset of peptides for which highly accurate quantification is required or the peptide is difficult to characterize by method (A), we will use a different targeted MS method on a Q-Exactive instrument (Thermo) available in the Proteomics Resource Center headed by Dr. Beatrix Ueberheide (NYU). The instrument is capable of a method called high-resolution/accurate mass acquisition which allows for comparatively straightforward, highly accurate fragment-ion based quantification. The method can also multiplex, i.e. measure several peptide precursor ions during one MS scan.

**Measurements.** We will focus on the stress vs. control condition (*without* the use of ubiquitination inhibitors) to monitor protein degradation, ubiquitination, and oxidation, and lack thereof. To monitor protein degradation, we will analyze three to five unmodified peptides per protein in their changes in ion intensity across the pulsed-SILAC samples and estimate degradation rates. To monitor ubiquitination and oxidation, we will search for mass-to-charge ratios for both the unmodified and modified predicted peptides in the standard SILAC samples. If peptides are not detected in the lysate, we may enrich for the specific protein by immunoprecipitation with either the antibody against the native protein or against the TAP tag. TAP-tagged strains will be purchased (Open Biosystems). To assess specificity and lack of ubiquitination, we will not only test peptides predicted to be the most likely ubiquitination sites, but also other lysine-containing peptides within the same protein that are predicted to *not* be ubiquitinated. The UB linkage type will be confirmed by western blotting after protein isolation. If K48 poly-ubiquitination occurs, the site and its ubiquitination are highly likely to be responsible for degradation. Due to their undefined sequence signature, oxidation events are more difficult to measure, but can be recognized as mass additions at some amino acids [89].

## 5. Broader impact

### Mentoring and training within the Vogel lab

**Ethnic and gender diversity.** Our research team is committed to diversity: Dr. Silva is Afro-American/Latino, and current undergraduate research students are of diverse ethnicity. We will continue to actively seek out and recruit scientists from under-represented minorities to participate in our research in our commitment to increase diversity in the research program. The PI also actively supports female students and scientists. The Vogel lab currently has six female lab members including the PI herself (post-doctoral researcher, graduate student, two master students, one undergraduate student) who the PI mentors for their respective career level, including career advice, presentation and interview help.

**Training in the Vogel lab.** The Vogel lab meets once a week to report on research progress and recent scientific literature, with all group members participating through presentations and discussions. This gives an opportunity for graduate students and postdoctoral research associates to practice and improve their oral communication skills. All graduate students are required to enroll in the NYU graduate biology course entitled “The Art of Scientific Investigation”, taught annually and which covers all aspects of professional development, including written and oral presentations, grantsmanship, and teaching. A graduate student for this project will be recruited from the NYU graduate program at the Center for Genomics and Systems Biology, while the post-doctoral researcher (Dr. Gustavo Silva) is already working in the Vogel lab. The graduate student will be jointly supervised by Drs. Vogel and Shasha (PI, CoPI) through weekly meetings and participation in respective group activities.

**Cross-disciplinary training.** To actively integrate molecular biology, computational sciences, and technology, the Vogel lab participates in several activities. The Center for Genomics and Systems Biology (which Dr. Vogel is part of) together with the NYU Courant Institute for Mathematical Sciences (which is arguably one of the best math institutions in the world) offer a variety of courses in which Dr. Silva and other lab members participate. In particular, these are e.g. G23.1128 Systems Biology; G23.1130 Applied Genomics & Modeling; G23.1127 Bioinformatics & Genomes. Dr. Silva has successfully completed the Statistics in Biology course (BIOL-GA.2030SP12, Dr. Tranchina). In addition, the Vogel lab is an active member of the New York Proteomics Special Interest Group that (across New York and New Jersey), with seminars, meetings, and discussions on issues related to mass spectrometry and quantitative proteomics. The Vogel lab is also actively discussing mathematical issues with Drs. Richard Bonneau and Dan Tranchina (both joint appointments between the Center for Genomics and Systems Biology and the Courant Institute), as well as Dr. Dennis Shasha (CoPI).

### LANG Program at the American Museum for Natural History

**Highschool student research experience**. To provide an opportunity to experience an interdisciplinary research laboratory to a larger group of students, Dr. Vogel is in contact with Drs. Noah Burg/Robert Habig from the American Museum of Natural History who organize the LANG program. The goal of this program is to increase the participation of minority children and children from low-income families in science. Dr. Vogel has committed to participating in lectures that are part of the LANG program (for 5th graders) and organizing a group visit to the Vogel lab once per year.

# References

1. Hastie, T., R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* 2001, New York: Springer.

2. Belle, A., A. Tanay, L. Bitincka, R. Shamir and E.K. O'Shea, *Quantification of protein half-lives in the budding yeast proteome.* Proc Natl Acad Sci U S A, 2006. **103**(35): p. 13004-9.

3. Eden, E., N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo and U. Alon, *Proteome half-life dynamics in living human cells.* Science, 2011. **331**(6018): p. 764-8.

4. Yen, H.C., Q. Xu, D.M. Chou, Z. Zhao and S.J. Elledge, *Global protein stability profiling in mammalian cells.* Science, 2008. **322**(5903): p. 918-23.

5. Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach, *Global quantification of mammalian gene expression control.* Nature, 2011.

6. Gulshan, K., B. Thommandru and W.S. Moye-Rowley, *Proteolytic degradation of the Yap1 transcription factor is regulated by subcellular localization and the E3 ubiquitin ligase Not4.* J Biol Chem, 2012. **287**(32): p. 26796-805.

7. Jung, T., A. Hohn, B. Catalgol and T. Grune, *Age-related differences in oxidative protein-damage in young and senescent fibroblasts.* Arch Biochem Biophys, 2009. **483**(1): p. 127-35.

8. Goldberg, A.L., *Protein degradation and protection against misfolded or damaged proteins.* Nature, 2003. **426**(6968): p. 895-9.

9. Pickart, C.M. and D. Fushman, *Polyubiquitin chains: polymeric protein signals.* Curr Opin Chem Biol, 2004. **8**(6): p. 610-6.

10. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu Rev Biochem, 1998. **67**: p. 425-79.

11. Shang, F., T.R. Nowell, Jr. and A. Taylor, *Removal of oxidatively damaged proteins from lens cells by the ubiquitin-proteasome pathway.* Exp Eye Res, 2001. **73**(2): p. 229-38.

12. Dudek, E.J., F. Shang, P. Valverde, Q. Liu, M. Hobbs and A. Taylor, *Selectivity of the ubiquitin pathway for oxidatively modified proteins: relevance to protein precipitation diseases.* FASEB J, 2005. **19**(12): p. 1707-9.

13. Medicherla, B. and A.L. Goldberg, *Heat shock and oxygen radicals stimulate ubiquitin-dependent degradation mainly of newly synthesized proteins.* J Cell Biol, 2008. **182**(4): p. 663-73.

14. Lee, B.H., M.J. Lee, S. Park, D.C. Oh, S. Elsasser, P.C. Chen, C. Gartner, N. Dimova, J. Hanna, S.P. Gygi, S.M. Wilson, R.W. King and D. Finley, *Enhancement of proteasome activity by a small-molecule inhibitor of USP14.* Nature, 2010. **467**(7312): p. 179-84.

15. Inai, Y. and M. Nishikimi, *Increased degradation of oxidized proteins in yeast defective in 26 S proteasome assembly.* Arch Biochem Biophys, 2002. **404**(2): p. 279-84.

16. Shringarpure, R., T. Grune, J. Mehlhase and K.J. Davies, *Ubiquitin conjugation is not required for the degradation of oxidized proteins by proteasome.* J Biol Chem, 2003. **278**(1): p. 311-8.

17. Asher, G., N. Reuven and Y. Shaul, *20S proteasomes and protein degradation "by default".* Bioessays, 2006. **28**(8): p. 844-9.

18. Kastle, M. and T. Grune, *Proteins bearing oxidation-induced carbonyl groups are not preferentially ubiquitinated.* Biochimie, 2011.

19. Li, W., M.H. Bengtson, A. Ulbrich, A. Matsuda, V.A. Reddy, A. Orth, S.K. Chanda, S. Batalov and C.A. Joazeiro, *Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling.* PLoS ONE, 2008. **3**(1): p. e1487.

20. Vierstra, R.D., *The ubiquitin-26S proteasome system at the nexus of plant biology.* Nat Rev Mol Cell Biol, 2009. **10**(6): p. 385-97.

21. Glickman, M.H. and A. Ciechanover, *The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction.* Physiol Rev, 2002. **82**(2): p. 373-428.

22. Wagner, S.A., P. Beli, B.T. Weinert, M.L. Nielsen, J. Cox, M. Mann and C. Choudhary, *A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles.* Mol Cell Proteomics, 2011.

23. Kim, W., E.J. Bennett, E.L. Huttlin, A. Guo, J. Li, A. Possemato, M.E. Sowa, R. Rad, J. Rush, M.J. Comb, J.W. Harper and S.P. Gygi, *Systematic and quantitative assessment of the ubiquitin-modified proteome.* Mol Cell, 2011. **44**(2): p. 325-40.

24. Xu, P., D.M. Duong, N.T. Seyfried, D. Cheng, Y. Xie, J. Robert, J. Rush, M. Hochstrasser, D. Finley and J. Peng, *Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation.* Cell, 2009. **137**(1): p. 133-45.

25. Sriram, S.M., B.Y. Kim and Y.T. Kwon, *The N-end rule pathway: emerging functions and molecular principles of substrate recognition.* Nat Rev Mol Cell Biol, 2011. **12**(11): p. 735-47.

26. Varshavsky, A., *The N-end rule pathway of protein degradation.* Genes Cells, 1997. **2**(1): p. 13-28.

27. Rechsteiner, M. and S.W. Rogers, *PEST sequences and regulation by proteolysis.* Trends Biochem Sci, 1996. **21**(7): p. 267-71.

28. Gsponer, J., M.E. Futschik, S.A. Teichmann and M.M. Babu, *Tight regulation of unstructured proteins: from transcript synthesis to protein degradation.* Science, 2008. **322**(5906): p. 1365-8.

29. Braun, B.C., M. Glickman, R. Kraft, B. Dahlmann, P.M. Kloetzel, D. Finley and M. Schmidt, *The base of the proteasome regulatory particle exhibits chaperone-like activity.* Nat Cell Biol, 1999. **1**(4): p. 221-6.

30. Strickland, E., K. Hakala, P.J. Thomas and G.N. DeMartino, *Recognition of misfolding proteins by PA700, the regulatory subcomplex of the 26 S proteasome.* J Biol Chem, 2000. **275**(8): p. 5565-72.

31. Bachmair, A., D. Finley and A. Varshavsky, *In vivo half-life of a protein is a function of its amino-terminal residue.* Science, 1986. **234**(4773): p. 179-86.

32. Gibbs, D.J., S.C. Lee, N.M. Isa, S. Gramuglia, T. Fukao, G.W. Bassel, C.S. Correia, F. Corbineau, F.L. Theodoulou, J. Bailey-Serres and M.J. Holdsworth, *Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants.* Nature, 2011. **479**(7373): p. 415-8.

33. Batonnet, S., M.P. Leibovitch, L. Tintignac and S.A. Leibovitch, *Critical role for lysine 133 in the nuclear ubiquitin-mediated degradation of MyoD.* J Biol Chem, 2004. **279**(7): p. 5413-20.

34. Jin, Y.J., C.Y. Cai, X. Zhang and S.J. Burakoff, *Lysine 144, a ubiquitin attachment site in HIV-1 Nef, is required for Nef-mediated CD4 down-regulation.* J Immunol, 2008. **180**(12): p. 7878-86.

35. Davies, M.J., *The oxidative environment and protein damage.* Biochim Biophys Acta, 2005. **1703**(2): p. 93-109.

36. Sohal, R.S., *Role of oxidative stress and protein oxidation in the aging process.* Free Radic Biol Med, 2002. **33**(1): p. 37-44.

37. Cabiscol, E., J. Tamarit and J. Ros, *Oxidative stress in bacteria and protein damage by reactive oxygen species.* Int Microbiol, 2000. **3**(1): p. 3-8.

38. Boschi-Muller, S., A. Gand and G. Branlant, *The methionine sulfoxide reductases: Catalysis and substrate specificities.* Arch Biochem Biophys, 2008. **474**(2): p. 266-73.

39. Biteau, B., J. Labarre and M.B. Toledano, *ATP-dependent reduction of cysteine-sulphinic acid by S. cerevisiae sulphiredoxin.* Nature, 2003. **425**(6961): p. 980-4.

40. Maisonneuve, E., B. Ezraty and S. Dukan, *Protein aggregates: an aging factor involved in cell death.* J Bacteriol, 2008. **190**(18): p. 6070-5.

41. Moller, I.M., P.E. Jensen and A. Hansson, *Oxidative modifications to cellular components in plants.* Annu Rev Plant Biol, 2007. **58**: p. 459-81.

42. Job, C., L. Rajjou, Y. Lovigny, M. Belghazi and D. Job, *Patterns of protein oxidation in Arabidopsis seeds and during germination.* Plant Physiol, 2005. **138**(2): p. 790-802.

43. Mariani, E., M.C. Polidori, A. Cherubini and P. Mecocci, *Oxidative stress in brain aging, neurodegenerative and vascular diseases: an overview.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **827**(1): p. 65-75.

44. Sayre, L.M., M.A. Smith and G. Perry, *Chemistry and biochemistry of oxidative stress in neurodegenerative disease.* Curr Med Chem, 2001. **8**(7): p. 721-38.

45. Jariel-Encontre, I., G. Bossis and M. Piechaczyk, *Ubiquitin-independent degradation of proteins by the proteasome.* Biochim Biophys Acta, 2008. **1786**(2): p. 153-77.

46. Tarcsa, E., G. Szymanska, S. Lecker, C.M. O'Connor and A.L. Goldberg, *Ca2+-free calmodulin and calmodulin damaged by in vitro aging are selectively degraded by 26 S proteasomes without ubiquitination.* J Biol Chem, 2000. **275**(27): p. 20295-301.

47. Orlowski, M. and S. Wilk, *Ubiquitin-independent proteolytic functions of the proteasome.* Arch Biochem Biophys, 2003. **415**(1): p. 1-5.

48. Pickering, A.M., A.L. Koop, C.Y. Teoh, G. Ermak, T. Grune and K.J. Davies, *The immunoproteasome, the 20S proteasome and the PA28alphabeta proteasome regulator are oxidative-stress-adaptive proteolytic complexes.* Biochem J, 2010. **432**(3): p. 585-94.

49. Grune, T., R. Shringarpure, N. Sitte and K. Davies, *Age-related changes in protein oxidation and proteolysis in mammalian cells.* J Gerontol A Biol Sci Med Sci, 2001. **56**(11): p. B459-67.

50. Yamanaka, K., H. Ishikawa, Y. Megumi, F. Tokunaga, M. Kanie, T.A. Rouault, I. Morishima, N. Minato, K. Ishimori and K. Iwai, *Identification of the ubiquitin-protein ligase that recognizes oxidized IRP2.* Nat Cell Biol, 2003. **5**(4): p. 336-40.

51. Vogel, C. and E.M. Marcotte, *Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data.* Nat Protoc, 2008. **3**(9): p. 1444-51.

52. Lu, P., C. Vogel, R. Wang, X. Yao and E.M. Marcotte, *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.* Nat Biotechnol, 2007. **25**(1): p. 117-24.

53. Vogel, C. and E.M. Marcotte, *Label-free protein quantitation using weighted spectral counting.* Methods Mol Biol, 2012. **893**: p. 321-41.

54. Vogel, C., G.M. Silva and E.M. Marcotte, *Protein expression regulation under oxidative stress.* Mol Cell Proteomics, 2011.

55. Chothia, C., J. Gough, C. Vogel and S.A. Teichmann, *Evolution of the protein repertoire.* Science, 2003. **300**(5626): p. 1701-3.

56. Vogel, C., C. Berzuini, M. Bashton, J. Gough and S.A. Teichmann, *Supra-domains - evolutionary units larger than single protein domains.* J Mol Biol, 2004. **336**(3): p. 809-23.

57. Vogel, C. and C. Chothia, *Protein family expansions and biological complexity.* PLoS Comput Biol, 2006. **2**(5): p. e48.

58. Vogel, C. and V. Morea, *Duplication, divergence and formation of novel protein topologies.* Bioessays, 2006. **28**(10): p. 973-8.

59. Vogel, C., S.A. Teichmann and J.B. Pereira-Leal, *The relationship between domain duplication and recombination.* J Mol Biol, 2005. **346**(1): p. 355-365.

60. Vogel, C., S. Abreu Rde, D. Ko, S.Y. Le, B.A. Shapiro, S.C. Burns, D. Sandhu, D.R. Boutz, E.M. Marcotte and L.O. Penalva, *Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line.* Mol Syst Biol, 2010. **6**: p. 400.

61. Kwon, T., H. Choi, C. Vogel, A.I. Nesvizhskii and E.M. Marcotte, *MSblender: a probabilistic approach for integrating peptide identifications from multiple database search engines.* J Proteome Res, 2011.

62. Ramakrishnan, S.R., C. Vogel, T. Kwon, L.O. Penalva, E.M. Marcotte and D.P. Miranker, *Mining gene functional networks to improve mass-spectrometry based protein identification.* Bioinformatics, 2009. **Jul 29**.

63. Ramakrishnan, S.R., C. Vogel, J.T. Prince, Z. Li, L.O. Penalva, M. Myers, E.M. Marcotte, D.P. Miranker and R. Wang, *Integrating shotgun proteomics and mRNA expression data to improve protein identification.* Bioinformatics, 2009. **25**(11): p. 1397-403.

64. Katari, M.S., S.D. Nowicki, F.F. Aceituno, D. Nero, J. Kelfer, L.P. Thompson, J.M. Cabello, R.S. Davidson, A.P. Goldberg, D.E. Shasha, G.M. Coruzzi and R.A. Gutierrez, *VirtualPlant: a software platform to support systems biology research.* Plant Physiol. **152**(2): p. 500-15.

65. Thum, K.E., M.J. Shin, R.A. Gutierrez, I. Mukherjee, M.S. Katari, D. Nero, D. Shasha and G.M. Coruzzi, *An integrated genetic, genomic and systems approach defines gene networks regulated by the interaction of light and carbon signaling pathways in Arabidopsis.* BMC Syst Biol, 2008. **2**: p. 31.

66. Gutierrez, R.A., L.V. Lejay, A. Dean, F. Chiaromonte, D.E. Shasha and G.M. Coruzzi, *Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis.* Genome Biol, 2007. **8**(1): p. R7.

67. Wang, J.T., H. Shan, D. Shasha and W.H. Piel, *Fast structural search in phylogenetic databases.* Evol Bioinform Online, 2005. **1**: p. 37-46.

68. Birnbaum, K., D.E. Shasha, J.Y. Wang, J.W. Jung, G.M. Lambert, D.W. Galbraith and P.N. Benfey, *A gene expression map of the Arabidopsis root.* Science, 2003. **302**(5652): p. 1956-60.

69. Shasha, D.E., *Plant systems biology: lessons from a fruitful collaboration.* Plant Physiol, 2003. **132**(2): p. 415-6.

70. Shasha, D.E., A.Y. Kouranov, L.V. Lejay, M.F. Chou and G.M. Coruzzi, *Using combinatorial design to study regulation by multiple input signals. A tool for parsimony in the post-genomics era.* Plant Physiol, 2001. **127**(4): p. 1590-4.

71. Birnbaum, K., P.N. Benfey and D.E. Shasha, *cis element/transcription factor analysis (cis/TF): a method for discovering transcription factor/cis element relationships.* Genome Res, 2001. **11**(9): p. 1567-73.

72. Silva, G.M., L.E. Netto, K.F. Discola, G.M. Piassa-Filho, D.C. Pimenta, J.A. Barcena and M. Demasi, *Role of glutaredoxin 2 and cytosolic thioredoxins in cysteinyl-based redox modification of the 20S proteasome.* FEBS J, 2008. **275**(11): p. 2942-55.

73. Silva, G.M., L.E. Netto, V. Simoes, L.F. Santos, F.C. Gozzo, M.A. Demasi, C.L. Oliveira, R.N. Bicev, C.F. Klitzke, M.C. Sogayar and M. Demasi, *Redox Control of 20S Proteasome Gating.* Antioxid Redox Signal, 2012. **16**(11): p. 1183-94.

74. Laurent, J., C. Vogel, T. Kwon, S. Craig, D.R. Boutz, H. Huse, K. Nozue, H. Walia, M. Whiteley, P. Ronald and E.M. Marcotte, *Protein abundances are more conserved than mRNA abundances across diverse taxa.* Proteomics, 2010. **10**(23): p. 4209-12.

75. Verma, R., S. Chen, R. Feldman, D. Schieltz, J. Yates, J. Dohmen and R.J. Deshaies, *Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes.* Mol Biol Cell, 2000. **11**(10): p. 3425-39.

76. Bastos de Oliveira, F.M., M.R. Harris, P. Brazauskas, R.A. de Bruin and M.B. Smolka, *Linking DNA replication checkpoint to MBF cell-cycle transcription reveals a distinct class of G1/S genes.* EMBO J, 2012. **31**(7): p. 1798-810.

77. Pannunzio, V.G., H.I. Burgos, M. Alonso, J.R. Mattoon, E.H. Ramos and C.A. Stella, *A Simple Chemical Method for Rendering Wild-Type Yeast Permeable to Brefeldin A That Does Not Require the Presence of an erg6 Mutation.* J Biomed Biotechnol, 2004. **2004**(3): p. 150-155.

78. Liu, C., J. Apodaca, L.E. Davis and H. Rao, *Proteasome inhibition in wild-type yeast Saccharomyces cerevisiae cells.* Biotechniques, 2007. **42**(2): p. 158, 160, 162.

79. Cox, J., I. Matic, M. Hilger, N. Nagaraj, M. Selbach, J.V. Olsen and M. Mann, *A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.* Nat Protoc, 2009. **4**(5): p. 698-705.

80. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

81. Cox, J., N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen and M. Mann, *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.* J Proteome Res, 2011. **10**(4): p. 1794-1805.

82. Flicek, P., B.L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S.C. Dyer, T. Eyre, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, R. Holland, K.L. Howe, K. Howe, N. Johnson, A. Jenkinson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A.J. Vilella, J. Vogel, S. White, M. Wood, E. Birney, T. Cox, V. Curwen, R. Durbin, X.M. Fernandez-Suarez, J. Herrero, T.J. Hubbard, A. Kasprzyk, G. Proctor, J. Smith, A. Ureta-Vidal, and S. Searle, *Ensembl 2008.* Nucleic Acids Res, 2008. **36**(Database issue): p. D707-14.

83. Keller, A., A.I. Nesvizhskii, E. Kolker and R. Aebersold, *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.

84. Nesvizhskii, A.I., A. Keller, E. Kolker and R. Aebersold, *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

85. Ong, S.E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey and M. Mann, *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

86. Brewis, I.A. and P. Brennan, *Proteomics technologies for the global identification and quantification of proteins.* Adv Protein Chem Struct Biol, 2010. **80**: p. 1-44.

87. Schwanhausser, B., M. Gossen, G. Dittmar and M. Selbach, *Global analysis of cellular protein translation by pulsed SILAC.* Proteomics, 2009. **9**(1): p. 205-9.

88. Lu, P., A. Rangan, S.Y. Chan, D.R. Appling, D.W. Hoffman and E.M. Marcotte, *Global metabolic changes following loss of a feedback loop reveal dynamic steady states of the yeast metabolome.* Metab Eng, 2007. **9**(1): p. 8-20.

89. Madian, A.G. and F.E. Regnier, *Proteomic identification of carbonylated proteins and their oxidation sites.* J Proteome Res, 2010. **9**(8): p. 3766-80.

90. Seyfried, N.T., P. Xu, D.M. Duong, D. Cheng, J. Hanfelt and J. Peng, *Systematic approach for validating the ubiquitinated proteome.* Anal Chem, 2008. **80**(11): p. 4161-9.

91. Mayor, T., J.R. Lipford, J. Graumann, G.T. Smith and R.J. Deshaies, *Analysis of polyubiquitin conjugates reveals that the Rpn10 substrate receptor contributes to the turnover of multiple proteasome targets.* Mol Cell Proteomics, 2005. **4**(6): p. 741-51.

92. Peng, J., D. Schwartz, J.E. Elias, C.C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley and S.P. Gygi, *A proteomics approach to understanding protein ubiquitination.* Nat Biotechnol, 2003. **21**(8): p. 921-6.

93. Starita, L.M., R.S. Lo, J.K. Eng, P.D. von Haller and S. Fields, *Sites of ubiquitin attachment in Saccharomyces cerevisiae.* Proteomics, 2011. **12**(2): p. 236-40.

94. Mirzaei, H. and F. Regnier, *Affinity chromatographic selection of carbonylated proteins followed by identification of oxidation sites using tandem mass spectrometry.* Anal Chem, 2005. **77**(8): p. 2386-92.

95. Mirzaei, H. and F. Regnier, *Identification of yeast oxidized proteins: chromatographic top-down approach for identification of carbonylated, fragmented and cross-linked proteins in yeast.* J Chromatogr A, 2007. **1141**(1): p. 22-31.

96. Hannay, K., E.M. Marcotte and C. Vogel, *Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation.* BMC Genomics, 2008. **9**: p. 609.

97. Wilson, D., M. Madera, C. Vogel, C. Chothia and J. Gough, *The SUPERFAMILY database in 2007: families and functions.* Nucleic Acids Res, 2007. **35**(Database issue): p. D308-13.

98. Talavera, D., C. Vogel, M. Orozco, S.A. Teichmann and X. de la Cruz, *The (In)dependence of Alternative Splicing and Gene Duplication.* PLoS Comput Biol, 2007. **3**(3): p. e33.

99. Vogel, C., S.A. Teichmann and C. Chothia, *Looking at the bigger picture.* Development, 2004. **131**(10): p. 2238-40.

100. Madera, M., C. Vogel, S.K. Kummerfeld, C. Chothia and J. Gough, *The SUPERFAMILY database in 2004: additions and improvements.* Nucleic Acids Res, 2004. **32**(1): p. D235-9.

101. Vogel, C., S.A. Teichmann and C. Chothia, *The immunoglobulin superfamily in Drosophila melanogaster and Caenorhabditis elegans and the evolution of complexity.* Development, 2003. **130**(25): p. 6317-28.

102. Bankier, A.T., H.F. Spriggs, B. Fartmann, B.A. Konfortov, M. Madera, C. Vogel, S.A. Teichmann, A. Ivens and P.H. Dear, *Integrated mapping, chromosomal sequencing and sequence analysis of Cryptosporidium parvum.* Genome Res, 2003. **13**(8): p. 1787-99.

103. Rogers, S., R. Wells and M. Rechsteiner, *Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis.* Science, 1986. **234**(4774): p. 364-8.

104. Vandermarliere, E. and L. Martens, *Protein structure as a means to triage proposed post-translational modification sites.* Proteomics, 2012.

105. Linding, R., L.J. Jensen, F. Diella, P. Bork, T.J. Gibson and R.B. Russell, *Protein disorder prediction: implications for structural proteomics.* Structure, 2003. **11**(11): p. 1453-9.

106. Ward, J.J., L.J. McGuffin, K. Bryson, B.F. Buxton and D.T. Jones, *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-9.

107. Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A.F. Quinn, C. Rivoire, A. Sangrador-Vegas, J.D. Selengut, C.J. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P.D. Thomas, C.H. Wu, C. Yeats, and S.Y. Yong, *InterPro in 2011: new developments in the family and domain prediction database.* Nucleic Acids Res, 2011. **40**(Database issue): p. D306-12.

108. de Lima Morais, D.A., H. Fang, O.J. Rackham, D. Wilson, R. Pethica, C. Chothia and J. Gough, *SUPERFAMILY 1.75 including a domain-centric gene ontology method.* Nucleic Acids Res, 2011. **39**(Database issue): p. D427-34.

109. Pethica, R.B., M. Levitt and J. Gough, *Evolutionarily consistent families in SCOP: sequence, structure and function.* BMC Struct Biol, 2012. **12**: p. 27.

110. Meslamani, J., D. Rognan and E. Kellenberger, *sc-PDB: a database for identifying variations and multiplicity of 'druggable' binding sites in proteins.* Bioinformatics, 2011. **27**(9): p. 1324-6.

111. Joosten, R.P., T.A. te Beek, E. Krieger, M.L. Hekkelman, R.W. Hooft, R. Schneider, C. Sander and G. Vriend, *A series of PDB related databases for everyday needs.* Nucleic Acids Res, 2011. **39**(Database issue): p. D411-9.

112. Bjorck, A., *Numerical methods for least squares problems.* SIAM J. Sci. Computing, 1996.

113. Socher, R., C. Lin, A.Y. Ng and C. Manning, *Parsing natural scenes and natural language with recursive neural networks.* Proceedings of the Twenty-Eighth International Conference on Machine Learning, 2011.

114. Richardson, N.E. and B.J. Meakin, *The influence of temperature on the sorption of benzocaine by nylon 6 from aqueous cosolvents.* J Pharm Pharmacol, 1977. **29**(2): p. 86-8.

115. Herrero, E., J. Ros, G. Belli and E. Cabiscol, *Redox control and oxidative stress in yeast cells.* Biochim Biophys Acta, 2008. **1780**(11): p. 1217-35.

116. Stephen, D.W., S.L. Rivers and D.J. Jamieson, *The role of the YAP1 and YAP2 genes in the regulation of the adaptive oxidative stress responses of Saccharomyces cerevisiae.* Mol Microbiol, 1995. **16**(3): p. 415-23.

117. Krems, B., C. Charizanis and K.D. Entian, *The response regulator-like protein Pos9/Skn7 of Saccharomyces cerevisiae is involved in oxidative stress resistance.* Curr Genet, 1996. **29**(4): p. 327-34.

118. Schmitt, A.P. and K. McEntee, *Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 5777-82.

119. Jang, H.H., K.O. Lee, Y.H. Chi, B.G. Jung, S.K. Park, J.H. Park, J.R. Lee, S.S. Lee, J.C. Moon, J.W. Yun, Y.O. Choi, W.Y. Kim, J.S. Kang, G.W. Cheong, D.J. Yun, S.G. Rhee, M.J. Cho and S.Y. Lee, *Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function.* Cell, 2004. **117**(5): p. 625-35.

120. Longo, V.D., E.B. Gralla and J.S. Valentine, *Superoxide dismutase activity is essential for stationary phase survival in Saccharomyces cerevisiae. Mitochondrial production of toxic oxygen species in vivo.* J Biol Chem, 1996. **271**(21): p. 12275-80.

121. Pedrajas, J.R., E. Kosmidou, A. Miranda-Vizuete, J.A. Gustafsson, A.P. Wright and G. Spyrou, *Identification and functional characterization of a novel mitochondrial thioredoxin system in Saccharomyces cerevisiae.* J Biol Chem, 1999. **274**(10): p. 6366-73.

122. Inoue, Y., T. Matsuda, K. Sugiyama, S. Izawa and A. Kimura, *Genetic analysis of glutathione peroxidase in oxidative stress response of Saccharomyces cerevisiae.* J Biol Chem, 1999. **274**(38): p. 27002-9.

123. Luikenhuis, S., G. Perrone, I.W. Dawes and C.M. Grant, *The yeast Saccharomyces cerevisiae contains two glutaredoxin genes that are required for protection against reactive oxygen species.* Mol Biol Cell, 1998. **9**(5): p. 1081-91.

124. Grant, C.M., L.P. Collinson, J.H. Roe and I.W. Dawes, *Yeast glutathione reductase is required for protection against oxidative stress and is a target gene for yAP-1 transcriptional regulation.* Mol Microbiol, 1996. **21**(1): p. 171-9.

125. Morano, K.A., C.M. Grant and W.S. Moye-Rowley, *The response to heat shock and oxidative stress in Saccharomyces cerevisiae.* Genetics, 2012. **190**(4): p. 1157-95.

126. Unsworth, E.R., P. Jones and S.J. Hill, *The effect of thermodynamic data on computer model predictions of uranium speciation in natural water systems.* J Environ Monit, 2002. **4**(4): p. 528-32.

127. Moskovitz, J., B.S. Berlett, J.M. Poston and E.R. Stadtman, *The yeast peptide-methionine sulfoxide reductase functions as an antioxidant in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9585-9.

128. Ramotar, D., S.C. Popoff, E.B. Gralla and B. Demple, *Cellular role of yeast Apn1 apurinic endonuclease/3'-diesterase: repair of oxidative and alkylation DNA damage and control of spontaneous mutation.* Mol Cell Biol, 1991. **11**(9): p. 4537-44.

129. Cheng, L., R. Watt and P.W. Piper, *Polyubiquitin gene expression contributes to oxidative stress resistance in respiratory yeast (Saccharomyces cerevisiae).* Mol Gen Genet, 1994. **243**(3): p. 358-62.

130. Lisowsky, T., *A high copy number of yeast gamma-glutamylcysteine synthetase suppresses a nuclear mutation affecting mitochondrial translation.* Curr Genet, 1993. **23**(5-6): p. 408-13.

131. Tamai, K.T., E.B. Gralla, L.M. Ellerby, J.S. Valentine and D.J. Thiele, *Yeast and mammalian metallothioneins functionally substitute for yeast copper-zinc superoxide dismutase.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8013-7.

132. Culotta, V.C., W.R. Howard and X.F. Liu, *CRS5 encodes a metallothionein-like protein in Saccharomyces cerevisiae.* J Biol Chem, 1994. **269**(41): p. 25295-302.

133. Juhnke, H., B. Krems, P. Kotter and K.D. Entian, *Mutants that show increased sensitivity to hydrogen peroxide reveal an important role for the pentose phosphate pathway in protection of yeast against oxidative stress.* Mol Gen Genet, 1996. **252**(4): p. 456-64.

134. Lewis, J.G., R.P. Learmonth, P.V. Attfield and K. Watson, *Stress co-tolerance and trehalose content in baking strains of Saccharomyces cerevisiae.* J Ind Microbiol Biotechnol, 1997. **18**(1): p. 30-6.

135. Flattery-O'Brien, J.A. and I.W. Dawes, *Hydrogen peroxide causes RAD9-dependent cell cycle arrest in G2 in Saccharomyces cerevisiae whereas menadione causes G1 arrest independent of RAD9 function.* J Biol Chem, 1998. **273**(15): p. 8564-71.

136. Izawa, S., Y. Inoue and A. Kimura, *Importance of catalase in the adaptive response to hydrogen peroxide: analysis of acatalasaemic Saccharomyces cerevisiae.* Biochem J, 1996. **320 ( Pt 1)**: p. 61-7.

137. Killcoyne, S., J. Handcock, T. Robinson, E.W. Deutsch and J. Boyle, *Interfaces to PeptideAtlas: a case study of standard data access systems.* Brief Bioinform, 2012. **13**(5): p. 615-26.

138. Deutsch, E.W., *The PeptideAtlas Project.* Methods Mol Biol, 2009. **604**: p. 285-96.