# Project Summary

## 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 can both modify and measure 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 start accumulating 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 both Stochastic Gradient Descent and ‘Regression Trees’ algorithms, learns 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 algorithms 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

## 1. Objectives

**Overarching goal.** Our goal is to manipulate protein expression levels through changes in protein stability. We have designed a strategy to address this question by taking advantage of a highly focused and specific system that modifies and measures 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 in yeast under steady state conditions and in response to H2O2 treatment. To learn which of the changes in rates may be dependent on ubiquitination, we will repeat the experiments but inhibit global ubiquitination with PYR-41. **Aim 1B - Modification.** Using metabolic labeling,enrichment systems, and mass spectrometry, we will identify proteins whose specific ubiquitination patterns change in response to H2O2. Oxidation propensity for each protein will also be monitored as it potentially influences stability and ubiquitination. **Aim 2A - Features.** Building on a published database of homology-modeled protein structures, and computational data mining, we will compile a comprehensive list of protein sequence and structure features associated with possible protein modification and subsequent degradation. **Aim 2B - Model.** We will integrate the computational (**2A**) and experimental (**1A,B**) data into a model that, using both Stochastic Gradient Descent and Boosted ‘Regression Trees’ algorithms [1], learns 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. **Aim 2C - Validation.** Using the model from **aim 2B**, we will predict the 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 modifications. The prediction of protein degradation will be validated by targeted mass spectrometry.

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

**=> Solving a debate.** The different routes of protein degradation upon stress, i.e. ubiquitin-dependent or –independent pathways, 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.

**=> 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. We will identify regulatory signatures of protein modification and degradation, and directly *test* the predictive model via targeted assays.

### Integration into a future research program (full NSF proposal)

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**. (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. (iii) Once modification sites 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 decrease in sensitivity.

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. 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 N-terminus (containing the E3 ligase interaction domain) was deleted, rendered the protein more stable and the cells less sensitive to stress [6] – directly connecting degradation regulation and an ‘improved’ stress response. Our preliminary model already successfully predicts the degradation propensity of Yap1 confirmed by this study (**Prelim. Results**).

## 2. Background

**Protein degradation as a crucial part of gene expression regulation.** Proteasomal degradation accounts for >90% of the normal cellular protein turnover [7], but even more so, failure to degrade oxidatively damaged proteins has detrimental effects for any cell [8]. Under oxidative stress, the oxidation of primarily proline, arginine, lysine, threonine, glutamate, and aspartate commonly generates carbonyl groups [9-11]. Despite the existence of stress defense and a few protein repair mechanisms [12, 13], the majority of oxidatively damaged proteins have to be removed from the cell to re-establish cellular proteostasis. In addition, some proteins appear to be more sensitive to oxidative stress than others, but the specific oxidative propensity for each protein is unknown. Accumulation of oxidized proteins/aggregates causes cellular senescence, loss of replicative ability and cell death [14].

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

**Poly-ubiquitination at lysine 48 (K48) allows the cell to target proteins for degradation in a highly specific manner** [15], but requires strict regulation and cellular energy. Poly-ubiquitination is a frequent and evolutionarily conserved [16-18] modification with a variety of cellular functions (via different lysine linkages). 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 [19, 20], 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, and they are negatively charged [19-21]. Poly-ubiquitination sites can be impacted by the protein’s N-terminus [22] and tend to occur in structured regions [20]. However, large-scale ubiquitination studies have so far only examined steady-state conditions, identifying global sequence signatures of ubiquitination [19, 20]. Despite the use of inhibitors and other approaches [19], it is unclear which of these ubiquitination events trigger degradation and which do not, nor is it clear under which conditions these roles may change.

**In addition, ubiquitin-independent protein degradation pathways exist.** Under oxidative stress, ubiquitin-dependent degradation is a major pathway [23-26], but many proteins are also removed via ubiquitin-independent degradation [27-30]. 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.

**Manipulating protein degradation is difficult.** Since protein degradation is an important part of gene expression regulation, researchers have attempted its manipulation. However, these attempts have proven to be challenging. Some studies have attempted to predict protein stability from sequence and structure, but little is known about stability *changes* under perturbation, e.g. stress. Some amino acids are thought to destabilize a protein if at the N-terminus [22, 31]; also proline, glutamate, serine, and threonine stretches can cause degradation [32]. Proteins with many intrinsically unstructured regions also degraded more rapidly than other proteins [33]. Changing the N-terminal residue or sequence stretch has been successfully used to stabilize a protein in individual cases [34, 35]. 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 [36, 37], but due to the lack of clear sequence signals, the experiments have operated on a trial-and-error basis.

**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 collection of sequence and structure based protein features, we will computationally learn to predict a protein’s fate with respect to degradation. We will validate the model on a set of stress-related proteins which have not been characterized in the experiments.

## 3. Investigator qualifications and preliminary data

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

Dr. Christine Vogel (PI) has extensive expertise in quantitative proteomics [38-41], protein sequence and structure analysis [42-46], and computational data analysis [47-50]. Dr. Dennis Shasha (CoPI) is an expert in biological data mining using mathematical approaches [51-58]. Dr. Gustavo Silva(postdoctoral researcher) has published extensively on proteasome regulation under oxidative stress and the ubiquitin/oxidation proteomic system [37, 59, 60].

### Published work

**Quantifying proteins at large scale (*Nat. Biotech 2007,25(1):117*; *Nat. 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 [38, 39].

**Protein translation and degradation substantially affect gene expression (*Nat. Rev. Genetics 2012, 13(4):227; Mol. Sys. Bio. 2010,6:400; Prot. 2010,10(23):4209*).** We demonstrated that protein concentrations are highly conserved across species – and more so than the corresponding mRNA concentrations [61]. Further, using mass spectrometry and computational analysis, we showed for >1,000 human proteins and ~150 sequence/structure features that transcription can only explain a small fraction (27%) of the variation in protein concentrations, 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 Prot., 2011,10(12)*)*.*** Using quantitative mass spectrometry, we analyzed time series data from >1,700 yeast proteins [38, 39, 41] and demonstrated that proteins have very different dynamics from RNA during the oxidative stress response indicating extensive regulation at the protein level: protein concentrations change at a slower rate than RNA.

### Pilot studies

**Both ubiquitin-dependent and –independent protein degradation occurs under oxidative stress.** Both oxidized and ubiquitinated proteins accumulate immediately after H2O2 treatment. In permeabilized yeast, we inhibit ubiquitination and proteasomal degradation by use of PYR-41 and MG-132, respectively (**Fig. 3.1**). The majority of oxidized proteins are not removed if proteasome function is impaired (**Fig. 3.1**)[8]. our results indicate that degradation of a fraction of oxidized protein delayed if ubiquitination is prevented, contradicting some published work [27, 28, 30]. Oxidized proteins seem to be largely tagged by K48-linked polyubiquitin, indicating their ubiquitin-dependent degradation (**Fig. 3.2**). These results suggest that (i) both ubiquitin-dependent and –independent degradation occurs in response to oxidative stress, and (ii)ubiquitination as a tag for degradation (K48) increases during the recovery phaseand (iii) damaged proteins are degraded about 2 to 4hrs *recovery*. We carefully designed the proposed experiments based on these results to measure degradation rates and to detect ubiquitin/oxidation profiles.

**Mass spectrometry identified oxidized, ubiquitinated proteins, and global protein expression.** We employed APEX-based mass spectrometry to semi-quantitatively estimate abundances of ~1,200 proteins in the whole cell extract and to identify proteins that are oxidized, or ubiquitinated (*not shown*). We find that different proteins follow very different pathways of ubiquitination, oxidation, and degradation under stress. For example, stress related proteins are both ubiquitinated and oxidized, but increase in expression.

|  |
| --- |
| **Tab. 3.1. Preliminary model.** Most predictive sequence features |
|  | **Positive** | **Negative** |
| **Degradation** | Lys, Pro, PEST motif | Codon adaptation, Arg, Isoel. point |
| **Ubiquitination** | Codon adaptation | Disorder , Lys, Length |
| **Oxidation** | Codon adaptation, Ile | Ser, Disorder, Length |

**Modeling can identify predictive sequence features.** Despite the (currently) semi-quantitative nature of our datasets and the lack of site-specific information, a preliminary model using the stochastic gradient descent algorithm 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 for protein degradation. 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](*not shown*).

## 4. Research methods

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

### Biological system and supporting infrastructure

|  |
| --- |
| NEW – Gustavo? |
| **Fig. 4.1. Overview - Experiments.**  |

**Strains and growth conditions.** All experiments will be conducted with the *S. cerevisiae* SILAC strain GMS125 derived from RJD1171 (MATa his3Δ200 leu2-3,112 lys2-801 trp1Δ63 arg4::KanMX6 ura3-52RPT1FH::Ylplac211 (URA3)[62]) that has one proteasome component (Rpt1) tagged with the FLAG epitope and a poly-histidine tail. To enable efficient inhibitor uptake (see below), cells are grown under conditions that induce permeability of the plasma membrane, i.e. in Minimal Proline Dextrose (MPD) medium [63]. Cells are allowed to divide at least 6 times and are treated in log phase (OD600~0.4) in the presence or not of inhibitors. [63]. 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 inhibitors if specified for recovery.

**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.

**Stable Isotopic Labeling of Amino acids in Cell culture (SILAC) to quantify protein modifications.** UB and OX enrichment/immunoprecipitation steps (see below) modify sample constitution and introduce errors. 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 [64]. The SILAC technique is very well established with standard protocols [64, 65]. SILAC approach improve the accuracy of quantitation, reduce the experimental biases since samples are prepared and run together, and also facilitate the rejection of contaminants. Pulsed-SILAC is a variation of isotope labeling in which the disappearance of fully labeled proteins is monitored after exposure to a pulse of a different isotope. Pulsed-SILAC will be used to measure degradation rates.

**Sample preparation and analysis.** Protein samples will be prepared as described before [41]. Briefly, cells are lysed, and lysate will be reduced, cysteines alkylated with iodoacetic acid, digested with trypsin, cleaned by C18 filtering. 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 [47]. All MS/MS data will be mapped to protein sequences and quantified using established pipelines, including those established in our lab [38, 39, 66-71].

### 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, 47], 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. We will investigate if protein-specific degradation is dependent on ubiquitination or not – contributing to answer to a long-debated question.

**Estimating protein degradation with pulsed-SILAC (Fig. 4.1B).** We will derive and compare estimates of protein degradation *rates* for two conditions: H2O2 treated cells (*stress*) 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. Protein concentration changes will be estimated from ion intensities [5] using standard software [66, 68]. The degradation rate DEG represents the slope of a curve fitted through the data [72] (**Fig. 4.1B**). If DEG=*rate\_H2O2/rate\_control* >0, the protein degrades faster under stress, and vice versa.

**Validation.** Alternative methods are simpler but more invasive; e.g. we can inhibit translation with cycloheximide, such that a decrease in protein concentrations (estimated by label-free APEX-based proteomics [38, 73]) is purely due to degradation. Alternative degradation pathways might be considered (e.g. autophagy) and also the extent of proteasomal dependent degradation by using MG-132 as potent inhibitor. Western blots and targeted MS experiments will serve to validate expression changes.

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

**Goal and expected outcomes.** Using SILAC, 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.

**Quantifying ubiquitination and oxidation.** To quantify ubiquitination and oxidation, samples will be combined from SILAC experiments (*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 (*LifeSensors*) which comprises arrays of Tandem Ubiquitin Binding Entities. To isolate oxidized proteins, we will derivatize one of the most common forms of oxidation [74], carbonylation, 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. The principles of this approach are well-established [19, 20] and will provide protein-specific information on the modification.

**Validation**. Protein ubiquitination can be validated in gel digestion/MS of well-defined molecular weight ranges and through literature and databases [75-78](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 [79, 80] 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. 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. 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.

**Sequence features** will include: sequence length, percentage of secondary structures, percentage of hydrophilic and hydrophobic, buried and surface amino acid residues, the presence of PEST-rich or unstructured regions and specific N-terminal residues (which are degradation signals [31, 33, 47, 81]), percentage of oxidation-prone residues at the protein surface, or vicinity of putatively oxidized residues to lysines (which may be ubiquitinated). Amino acids (and sequence stretches) will also receive an evolutionary conservation score, exploiting the fact that functionally relevant sequences are likely conserved.

**Structural features** (helping prediction of post-translational modification sites [82]) will include: (i) the propensity of sequence stretches to assume specific secondary structures, e.g. alpha helices, beta strands, or coils predicted from sequence with DisEmbl and DisoPred[83, 84]; and (ii) location of amino acid residues within the structure as predicted from solved structures or homology models available in PDB, SCOP [85-87], InterPro, SUPERFAMILLY [88, 89] or the Proteome Folding Project developed by our colleague Richard Bonneau (NYU)[90].

**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.

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

**Goal and expected outcomes.** We will integrate the computational (**2A**) and experimental (**1A,B**) data into a model that, uses both Stochastic Gradient Descent and Boosted ‘Regression Trees’ [1] to 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. The algorithms are complementary in that regression trees are easier to interpret, but stochastic gradient descent handles feature interaction better. The results of the modeling will, for the first time, outline general principles of protein degradation and modification for the entire yeast genome.

**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 (**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. The output of the model, for a given sequence, will be an estimate of its fate by predicting the values for OX, UB, DEG. 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, 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.

**Combined effects.** We will also analyze (using both algorithms) the product terms of two feature values to identify possible *combined* effects and interdependencies between two features. For example, we may evaluate the relationship between protein length andthe degree of unstructuredness*.* 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 repeatedly 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**.

### 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 degradation will be validated by targeted mass spectrometry experiments which 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.

**Protein selection.** The validation experiments will focus on ~100 known stress-related proteins. While we expect to observe many of these proteins, the data may be incomplete (i.e. for a given protein, degradation may be measured, but not ubiquitination), unreliable (if only one replicate reports the protein), or absent (e.g. transcription factors of low abundance). 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 which i) have key roles in the stress protective pathway; ii) strongly prefer UB-dependent degradation as predicted by the model; iii) present a strong sequence signature (e.g. for ubiquitination); and iv) are measurable by MS as evidenced by MS data or APEX-predictions [38, 39]. We will include proteins degraded by both the UB-dependent and -independent pathways, to allow for positives and negatives for learning and predicting.

**Targeted mass spectrometry.** If many 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 [91, 92], we will prioritize their analysis based on predicted observability derived from the method we developed [38, 39]. 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, and quantify these peptides 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 (NYU Medical School). The instrument is capable of a method called high-resolution/accurate mass acquisition which allows for comparatively straightforward, highly accurate fragment-ion based quantification.

**Measurements and validation.** We will focus on the stress vs. control condition (*without* the use of ubiquitination inhibitors) to primarily 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. If peptides are not detected, we may enrich for the specific protein by immunoprecipitation with either the antibody against the native protein or against the TAP tagged proteins purchased individually. To assess specificity and lack of ubiquitination, we will validate a few selected proteins by western blotting. Due to their undefined sequence signature, oxidation events are more difficult to measure, but can be recognized as mass additions at some amino acids [74].

## 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. All graduate students are required to enroll in the NYU graduate biology course entitled “The Art of Scientific Investigation”, 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 and jointly supervised by Vogel and Shasha, while the post-doctoral researcher (Dr. Gustavo Silva) is already in the Vogel lab.

**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 together with the NYU Courant Institute for Mathematical Sciences offer a variety of courses in which Dr. Silva and other lab members participate. In particular, these are 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). The Vogel lab is an active member of the New York Proteomics Special Interest Group that, with seminars, meetings, and discussions on issues related to quantitative proteomics. The Vogel lab is also actively discussing mathematical issues with Drs. Richard Bonneau and Dan Tranchina, and Dr. Dennis Shasha (all NYU).

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

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

11. 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.

12. 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.

13. 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.

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

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

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

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

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

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

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

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

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

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

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

43. 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.

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

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

46. 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.

47. 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.

48. 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.

49. 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**.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

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

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

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

83. 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.

84. 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.

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

86. 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.

87. 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.

88. 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.

89. 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.

90. Drew, K., P. Winters, G.L. Butterfoss, V. Berstis, K. Uplinger, J. Armstrong, M. Riffle, E. Schweighofer, B. Bovermann, D.R. Goodlett, T.N. Davis, D. Shasha, L. Malmstrom and R. Bonneau, *The Proteome Folding Project: proteome-scale prediction of structure and function.* Genome Res, 2011. **21**(11): p. 1981-94.

91. 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.

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