Info: http://www.nsf.gov/pubs/policydocs/pappguide/nsf09\_1/gpg\_2.jsp#IID2

**Rational Protein Stabilization – Evaluation of Strategies and Targets**

**Premise.** Our goal is to manipulate protein stability through changes in specific sequence sites. One of the most efficient strategies to do so is to modify a ubiquitinated lysine (and therefore prevent ubiqutination) – however, as several different types of polyubiquitination can occur within one protein and not all of them lead to degradation, determination of degradation-specific ubiquitination events and their recognition motifs within the protein has been a very difficult task.

**Proposed work.** We have designed an alternative strategy to address this question by taking advantage of a highly focused and specific system that explores ubiquitin-dependence of protein degradation in response to oxidative stress. We benefit from the use of (i) the robust and mutation-amenable yeast as a model system; (ii) precisely timed oxidative stress treatment to maximize the prevalence of degradation-specific K48 polyubiquitination over other modifications; (iii) use of new enrichment systems which, in combination with inhibitors and large-scale proteomics, deliver key information on which protein follows which pathway with respect to poly-ubiquitination and degradation. Using this system, we will identify, for thousands of proteins, if degradation rates change in response to oxidative stress and if the proteins are K48 polyubiquitinated. 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 and ubiquitination based on its amino acid sequence. 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.

**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 (in general), we will repeat the experiments but inhibit global ubiquitination with PYR-41.

**Aim 1B - Modification.** Using metabolic labeling, a new K48-polyubiquitination-specific enrichment system (TUBE), and mass spectrometry, we will identify proteins whose K48-specific modification pattern changes in response to H2O2. Proteins that are found to be K48-polyubiquitinated (**1B**) and change degradation when ubiquitination is globally inhibited (**1A**) are strong candidates for ubiquitination-dependent protein degradation and primary targets for future manipulation.

**Aim 2A - Features.** Building on an extensive set of amino acid sequence features, a 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 an algorithm called Regression Trees, learns for each experimentally observed proteins which sequence and structure features are predictive of (i) its change in degradation rate under oxidative stress, and (ii) the ubiquitination 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 ubiquitination signals. Our preliminary results have shown that lysines 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 K48-polyubiquitination. If both are the case, the most likely K48-ubiquitination *site* will be identified with the help of the most predictive features identified in **aim 2B**. Both the degradation of the entire protein, and the ubiquitination of a specific lysine-containing peptide, 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.

**Transformative aspects.** *(CV: not sure we need this, but the EAGER description mentions that)* The proposed work provides the first comprehensive and quantitative evaluation of the ubiquitin-dependence of protein degradation under oxidative stress (which has been the subject of long-standing debate). We use recently developed molecular tools and state-of-the-art mass spectrometry techniques to conduct both highly specific, but also genome-wide and highly representative large-scale experiments. In extensive pilot studies, we have identified a system which allows us to trace K48-specific ubiquitination with minimal contamination by other ubiquitination types and modifications. Finally, 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 degradation and ubiquitination behavior under stress.

**Broader context.** *(CV: possibly only for us to see how it fits into context)* The proposed work provides key building blocks for the next steps in understanding protein degradation upon oxidative stress. (i) With highly sensitive, but experimentally non-trivial *site-specific* ubiquitination assays, we can extract *local* information on likely ubiquitination sites and their sequence contexts. Given that the average protein contains ~20 lysines (all of which can be ubiquitinated), the method will refine the predictions based on the data from **aim 1B** above. (ii) Another factor confounding polyubiquitination of lysines is *oxidation* of proteins that provokes structural changes. A set of experiments can quantify the extent of oxidation and provide data to include oxidation propensity of proteins into the predictive model (**2B**). (iii) Once K48-specific polyubiquitination sites have been identified and validated, the next step is to manipulate degradation behavior of the respective protein through site-specific mutagenesis. Such experiments, when conducted on stress-related proteins, can be evaluated in the protective effect on the cell. [I don’t understand this last phrase. Do mean “can determine whether such manipulations can protect a cell”?]