

# Degradome project: Insights into intracellular protein degradation

Guy E. Zinman<sup>1,2\*</sup>, Thomas S. Jensen<sup>2</sup>, Søren Brunak<sup>2</sup>

1. Computational Biology department, Carnegie Mellon University, Pittsburgh, USA

2. Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark

\*email: zinman@cmu.edu

## Introduction

Protein molecules are continuously synthesized and degraded in the cell, and the balance between the two competing processes is important for the proper function of the cell. Active protein degradation also plays important role in many key cellular pathways like cell cycle progression, apoptosis, signal transduction and cell development. Moreover, misfolded proteins and incompletely synthesized proteins should be removed to ensure intact operation of cellular processes.

The ubiquitin system is identified as the main pathway for labeling targeted proteins to destruction. Despite intensive research, no sequence motif was identified as a recognition site for the ubiquitination pathway enzymes or the ubiquitin molecule destruction signal. The known genetic and biochemical methods for identifying actively degraded proteins currently cannot be performed large scale. Hence, a novel approach was developed for characterizing actively degraded proteins and selecting the most promising and interesting protein candidates to be tested.

## Method

The Degradome project is a novel in-silico computational approach aimed at characterizing actively degraded proteins by their physical properties and features, and complementing the experimental results in establishing the correct set of ubiquitin-conjugated protein targets and actively intracellular degraded proteins.

In this study, several feature-based artificial neural networks (Figure 1) were constructed in order to predict potential ubiquitinated substrates and short-lived proteins. The predictors were trained on the yeast proteome and tested on the yeast and human proteomes. Features like amino acid composition, negatively charged atoms, sequence length, localization, were found to be important for distinguishing ubiquitinated proteins as well as short-lived proteins. Diverse proteins from distinct compartments and pathways are being actively degraded; hence the ability to describe these proteins using few sequence-derived features is notable by itself.

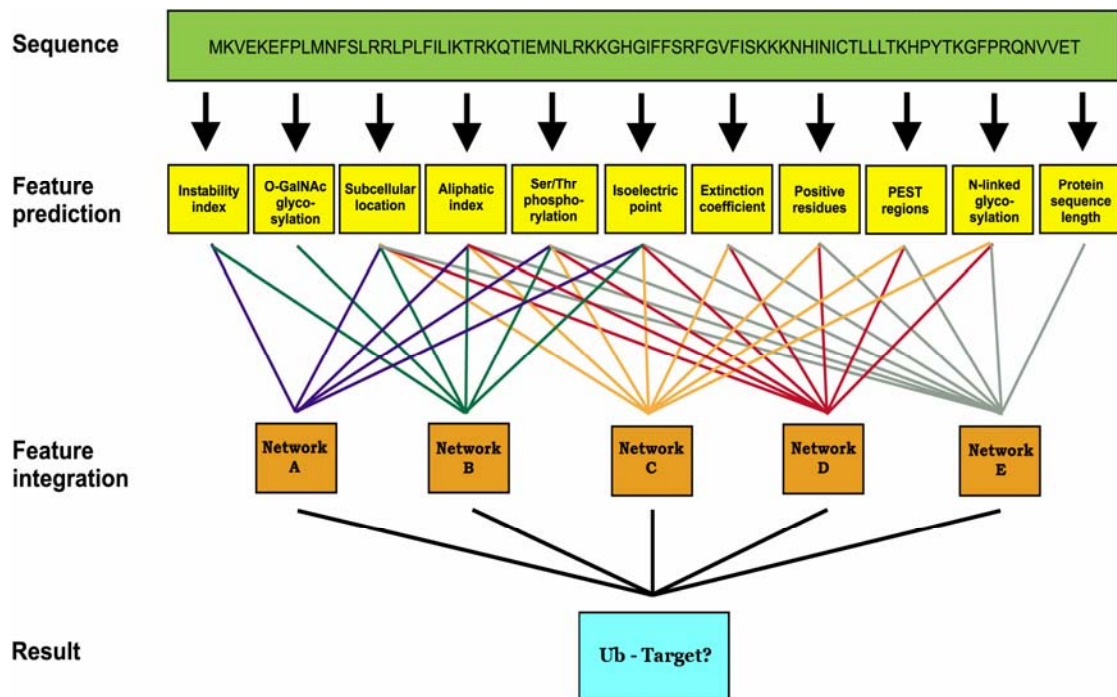


Figure 1: For every protein we derive features (yellow boxes) directly from the amino acid sequence using a set of well documented tools and prediction servers. Proteins are thus represented by their feature vector. Each neural network in the ensemble integrates a different set of features, and the output from all networks is combined into one final score, expressing to what degree the protein possesses features characteristic of ubiquitin target or short-lived protein.

## Results

The predictors for identifying potential ubiquitination targets and proteins with short half-life obtained a reasonable correlation coefficient close to 0.5 on the yeast test dataset. Testing the ubiquitin predictor on the human proteome achieved a correlation of 0.32, predicting 64% of the ubiquitinated proteins in human correctly. Many relevant over-represented GO terms were found in the analyses of all the predictors, enhancing the confidence in the ability of the predictors to correctly identify actively degraded proteins.

To conclude, the Degradome project is a novel tool for predicting candidates for ubiquitination and active degradation, which is far superior to other methods presently used for getting degradation clues.