Supporting Methods

Quantification of Protein Half-Lives in the Budding Yeast Proteome

1. Cell Growth and Cycloheximide Treatment

Three parallel cultures (1.7 mL) of each TAP-tagged strain were grown in separate 96-well plates in YEPD medium to log phase (OD_{600} ~ 0.5) at 30^0C. Cycloheximide (Sigma), a translation inhibitor, was added to a final concentration of 35 μg/mL to terminate protein synthesis (1,2). This concentration of cycloheximide was selected to be high enough to inhibit protein synthesis without causing significant growth defects during the course of the experiment (data not shown). We observe a ~10% decrease in cell growth 30 min following cycloheximide treatment. Following cycloheximide treatment, equal numbers of cells (as determined by OD_{600} units) were collected at 0, 15 and 45 min by centrifugation at 4^0C and flash-frozen in liquid nitrogen.

2. Extract Preparation and Western Blot analysis

Cell lysates were prepared as described in ref. 3, with the following minor modifications. Pelleted cells were lysed by addition of 60 µL of hot SDS-PAGE sample buffer followed by boiling for 5 min. Lysed cells were centrifuged, the supernatant extracts were reordered into fresh 96-well plates such that the three time points corresponding to a given TAP-tagged strain were located adjacent to each other, and lysates were frozen at -80^0C. 14 µL aliquots of the lysates were loaded onto 26 well, pre-cast 4-15% gradient acrylamide Tris-HCl Criterion gels (Bio-Rad) such that the three time points corresponding to a given TAP-tagged strain were run in adjacent lanes of the same gel, thereby avoiding errors introduced by gel-to-gel variability in the Western analysis. A pre-stained molecular weight ladder (Bio Rad) and “Magic Mark” (Invitrogen) molecular weight standards were included on each gel. The gels were run at 150 Volts for 100 min and transferred onto nitrocellulose membranes using the Trans-Blot cell (Bio Rad) at 1200 mA for 90 min in 20 mM $NaPO_4$, pH 6.8 buffer. The blots were probed using affinity purified rabbit polyclonal antibody raised against the calmodulin binding peptide (1:5,000 dilution; Open Biosystems). This antibody detects the TAP tag with great sensitivity as it can bind CBP as well as the protein A segment of the tag. The blots were subsequently probed with horseradish peroxidase (HRP) conjugated goat secondary antibody (1:20,000 dilution; Jackson ImmunoResearch) against rabbit IgG and developed with SuperSignal West Femto Maximum Sensitivity Substrate ECL (Bio Rad). The chemiluminescent signals of the bands corresponding to the TAP-tagged proteins were detected at three exposure times (30 sec, 1 min. and 5 min) using a CCD camera (FluorChem 8800, Alpha Innotech).

3. Image Analysis and Data Generation

The image analysis is divided into three phases: (*i*) protein band definition, (*ii*) gel exposure selection, and (*iii*) signal intensity evaluation. During phase *(i)* the custom software (QuantiAction) works together with the user to define bands.

First, the user selects an area of the gel representing a complete time course for one protein; this selection consists of a rectangular box enclosing all the bands. Second, the software finds the lane separation points within the user's selection, thus defining inner rectangular boxes, each enclosing only one band. Multiple exposure times (10 sec, 1 min and 5 min) of gel images were acquired using a CCD camera in 16-bit gray scale levels. Longer exposure times were used to enhance the signal to noise ratio of low abundance proteins while shorter exposure times ensure that the signal of abundant proteins does not reach saturation.

During phase (*ii*), for each protein we select the longest exposure in which none of the bands in the time course has reached saturation and use this for half-life calculations. We define saturation as the number of pixels within a band whose value is on or above the maximum empirical pixel value that the camera can generate.

During phase (*iii*) the band selection information is used on the exposure chosen during phase (*ii*) to measure the band signals. Background correction is estimated based on the assumption that at least one of the upper or lower lines of the band's bounding box lie in the background. The background signal level is calculated as the mean of the pixels in both (if upper and lower have similar pixel values) or just one (the line with lower pixel values) of the upper and lower lines of the bounding box. The background level is subtracted from each pixel inside the bounding box. The band's signal is then defined as the sum of all pixels inside the bounding box whose value is greater than 0. A standard linear least squares fit method is then used to estimate the degradation rate constant. Finally, all the data entered by the user or generated by the software, including the band selection box coordinates, is then saved in the database for later retrieval and analysis.

4. First-Order Decay Kinetics of Protein Degradation

To determine if, *on average,* protein degradation follows first order decay kinetics, we performed the following analysis:

1. Normalize the background corrected intensity of each time course with respect to the intensity of time-point 0 via the following formula: $I_n(t) = \frac{I(t)}{I(0)}$. This normalization is necessary when comparing intensities measured on different Western blots.
2. Average the normalized intensities for each time point over all protein time-courses using the following formula: $\overline{I_n(t)} = \frac{1}{N}\sum_{i=1}^{N} \ln\left[I_n^i(t)\right]$, where *N* is the number of proteins for which we have obtained measurements and $I_n^i(t)$ is the normalized intensity of the i protein.
3. Fit a first order exponential by using the linear least square fit method on the averaged log-transformed normalized intensities, $\overline{I_n(t)}$,.

4. Determine the significance of the correlation coefficient of the fitted exponential.

We observed Pearson R = -0.96, P < 0.09. The significance of the correlation is influenced (negatively) by the small number of data points used to fit the line however, this correlation supports, at the 10% significance level, the hypothesis that, on average, the protein degradation follows first order decay kinetics.

5. Measurement Error Analysis

We assumed that our half life estimates are the product of two variables: (i) protein half-life ($T_{1/2}$) and (ii) a log-normally distributed multiplicative (additive in log scale) error factor (ε) with mean, $\mu_{\ln(\varepsilon)} = 0$ and unknown variance. Given a random set of replicate half life measurements (T_1, T_2), we show that the multiplicative error factor is log-normally distributed by plotting the distribution of $\ln(T_1/T_2)$ and observing that it is approximately normal (Fig. 6) and is only weakly dependent in the absolute value of T_1 or T_2 (data not shown). We estimate the variance, in log scale, of the multiplicative error term, ln(ε) as half the variance of the random variable $\ln(T_1/T_2)$. The variance of the error term reflects the anticipated error in our measurements. To compare this internal experimental error to systematic errors resulting from using TAP tags or from the use of cycloheximide, we computed Kolmogorov-Smirnov (KS) statistics on the distributions of differences between replicate experiments and the distribution of differences between our measurements and literature half lives or half lives derived using specific antibodies.

6. Dilution Series

To estimate the dynamic range and the linearity limit of our Western blotting technique, we carried out dilution experiments. Using purified protein, TAP-tagged *Escherichia coli* initiation factor A (INFA) (3), we performed serial dilutions with dilution factor $X \in \{0.5,\ 0.25,\ 0.125\}$ where $X = 1 - \frac{P_t}{P_{t-1}}$. Here we describe our results and more details for X = 0.125.

Note that this experiment is different from other controls because purified protein (spiked into wild type yeast extract) is used and the errors involved in protein extraction steps are eliminated. This experiment is primarily designed to measure the sensitivity and resolution of Western blots used for quantitative intensity measurement.

We start with 6.6234 fmol of INFA and do $\frac{7}{8}$ dilutions in 21 sequential steps down to 0.4011 fmol. We use the following formula for converting from fmol to molecules/cell:

$$1.7\,\cancel{mL} \times 4\times10^{7}\,\frac{cells}{\cancel{mL}\times\cancel{OD}} \times 0.5\,\cancel{OD} \times \frac{14\,\cancel{mL}}{60\,\cancel{mL}} \times 6.023\times10^{-23}\,\frac{mole}{molec} = 1.31717\times10^{-17}\,\frac{cells\times mole}{molec}$$

$$1\,fmole = 75\,\frac{molucules}{cell}$$

Given that we could detect 0.4011 fmol, and using the above formula, the dilution range translates into 30-500 molecules/cell, or the least abundant 14% of the yeast proteome for which a valid abundance measurement is available. Note that the lowest reported abundance is ~50 molecules per cell (3).

To determine whether our measurement technique can detect the $\frac{7}{8}$ dilution, we analyzed the background-corrected intensity measurements generated by QuantiAction. We compared the observed and the input dilution factors. To estimate the observed dilution factor we initially log transformed the background-corrected intensity measurements and then used linear least squares fit to estimate the correlation between log protein intensity and dilution number. We observed a correlation R = 0.9822, P < 2.9e-15, the observed dilution factor was 0.1184 with standard error 0.00385 (Table 9 and Fig. 8). Since the 95% confidence interval of the correlation (0.1108 – 0.1262) contains the input dilution factor, 0.125, we conclude that our measurement technique can reliably detect a $\frac{7}{8}$ signal change. Given that our most reliable data points are separated by 15 min (time point 0 and 15 min), using these two points alone to calculate half-life, this signal change would translate into a half-life of approximately 78 minutes. Therefore, 78 minutes would be the maximum detectable half-life under ideal conditions; however given that in our experiment there is an extra step (of cell lysis) before the Western procedure, we conservatively cap the half-lives at 60 min.

7. Clustering

We combined information on three parameters of proteins: (*i*) protein production rate (mRNA abundance x ribosome density), (*ii*) protein abundance and (*iii*) protein half-life. We ranked the proteins according to each of the three parameters and assigned each protein a score: positive, if they are in the upper 40 percentile for a given parameter; negative if they are in the lower 40 percentile; or undefined if they are in the middle 20 percentile. Proteins that were classified either as positive or negative for all three parameters were grouped into clusters according to their behavior for the three parameters. This analysis yielded eight disjoint cluster configurations (+++, ++-, +--, ---, etc); two of these clusters contain the majority of the proteins (++- and --+). We then computed GO functional enrichments for each cluster using the TANGO program (4) fully correcting for multiple testing.

8. Physical Attributes Analysis

To determine whether there exists any relationship between protein half-life and primary sequence properties, we correlated the half-life versus all amino acid densities. The density of each amino acid is computed as the number of that specific amino acid divided by the length of the protein. This analysis reveals a coprrelation between higher densities of serine and shorter half lives, while higher densities of valine correlate with longer half-lives (Table 6 and Fig.

7A). We also observed a significant negative correlation between protein half-life and protein length (Spearman *R* = -0.23, *P* < 3e-38, Fig. 7B).

9. Protein Metabolism at Steady State

To visualized the dependencies among protein abundance, production, and degradation (Fig. 5) we used a bin size of 0.64 for the log production rate and 1.6 for the log degradation rate.

The rate of change in protein concentration can be modeled using the following differential equation:

$$\dot{P}(t) = \frac{dP(t)}{dt} = M(t) * R - P(t) * (D + V) \qquad (1)$$

where *P* is the protein concentration, *M* is the absolute mRNA concentration, *R* is the rate of translation per mRNA molecule (corresponding to the ribosome density), *D* is the protein degradation rate constant, and *V* is the growth rate (volume increase factor per unit time). At steady state the protein concentration is constant over time or, mathematically, at *t=0*:

$$\begin{aligned} \dot{P}(0) &= M(0) * R - P(0) * (D + V) \\ P(0) &= \frac{M(0) * R}{D + V} \qquad (2) \\ \log(P(0)) &= \log(M(0)) + \log(R) - \log(D + V) \end{aligned}$$

To test the compatibility of the various data sources and their adequacy for quantitative modeling we used experimentally determined data for *P, M, R* and *D*. We set $V = 2^{\frac{1}{90}} = 0.0077$, since the doubling time is approximately 90 min. We transformed all data sources to a log scale and computed the Spearman and Pearson correlation coefficient for log *(P) vs. a)* log *(M)* + log*(R)* and *b)* log *(M)* + log*(R)* – log *(D + V).* For this analysis the *M* and *R-values* were obtained from Beyer *et al.* (that data set gave slightly more accurate results than the other datasets). The *V* parameter *(the growth rate)* was important for obtaining a valid correlation and dominated *D* for stable proteins.

To test the significance of the contribution of the degradation rate constant to protein abundance determination we performed residual analysis (5) using the following equation:

$$G = \log(P) - \left(a * \left[\log(M) + \log(R)\right] + b\right)$$

where log (*P*) is the measured protein abundance and $\left(a * \left[\log(M) + \log(R)\right] + b\right)$ is the predicted protein abundance as defined without using the degradation term. We then computed the Spearman correlation between *G* and log *(D + V).*

10. A Model Relating Protein Half-Life and Transcriptional Control Within Groups of Co-regulated Genes

As described in the main text, we are analyzing the transcriptional regulation of a set of co-regulated genes, assuming the cell tries to generate a coordinated response at the protein level by balancing differences in the

degradation rates using variable transcriptional control. If $\pi(t)$ represents the common fold change in protein concentrations for all the proteins in a co-regulated module, we derive the following constraint on the protein abundances:

$$\frac{\dot{P(t)}}{P(t)} = \frac{\left(P(0) * \dot{\pi(t)}\right)}{P(t)} = \frac{P(0) * \dot{\pi(t)}}{P(t)} = \frac{\dot{\pi(t)}}{\pi(t)} \quad (3)$$

We will use equations 1 and 2 to transform this constraint into a relationship between *M (t)* (expression) and *D* (protein degradation). To characterize this relationship, we assume that the translation rate and degradation rate constant for each protein are not regulated and remain constant throughout the response to a new condition or perturbation. We first use the eq. 1 to rewrite eq. 3 in terms of *M, D* and *R,* as follows:

$$\frac{\dot{P(t)}}{P(t)} = \frac{M(t) * R - P(t)(D+V)}{P(t)}$$

Then substituting $P(t) = P(0) * \pi(t)$ we have:

$$\frac{\dot{P(t)}}{P(t)} = \frac{M(t) * R}{P(0)\pi(t)} - (D+V)$$

We next use eq. 2 to substitute $P(0) = \frac{M(0) * R}{D+V}$ and obtain:

$$\frac{\dot{P(t)}}{P(t)} = \frac{(D+V) * M(t) * \cancel{R}}{M(0) * \pi(t) * \cancel{R}} - (D+V) = (D+V)\left(\frac{M(t)}{M(0) * \pi(t)} - 1\right)$$

Using equation 3 we can substitute $\frac{\dot{P(t)}}{P(t)}$ with $\frac{\dot{\pi(t)}}{\pi(t)}$, and therefore write the expression fold change $\frac{M(t)}{M(0)}$ in terms of $\pi(t)$ and *D* as follows:

$$\frac{M(t)}{M(0)} = \frac{\dot{\pi(t)}}{(D+V)} + \pi(t) \quad (4)$$

As discussed in the main text, this equation predicts the correlation between expression fold change and protein degradation rates, suggesting negative and positive correlations for induced and repressed modules, respectively.

REFERENCES

1. Rao, S. S. & Grollman, A. P. (1967) *Biochem Biophys Res Commun* **29,** 696-704.
2. Zhou, P. (2004) *Methods Mol Biol* **284,** 67-77.
3. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. (2003) *Nature* **425,** 737-41.
4. Shamir, R., Maron-Katz, A., Tanay, A., Linhart, C., Steinfeld, I., Sharan, R., Shiloh, Y. & Elkon, R. (2005) *BMC Bioinformatics* **6,** 232.
5. Chen, P. Y. & Popovich, P. M. (2002) *Correlation: Parametric and Nonparametric Measures* (Sage Publications, Thousands Oaks, CA).