Global Protein Stability Profiling in Mammalian Cells

Hsueh-Chi Sherry Yen, Qikai Xu,* Danny M. Chou,* Zhenming Zhao, Stephen J. Elledge†

The abundance of cellular proteins is determined largely by the rate of transcription and translation coupled with the stability of individual proteins. Although we know a great deal about global transcript abundance, little is known about global protein stability. We present a highly parallel multiplexing strategy to monitor protein turnover on a global scale by coupling flow cytometry with microarray technology to track the stability of individual proteins within a complex mixture. We demonstrated the feasibility of this approach by measuring the stability of ~8000 human proteins and identifying proteasome substrates. The technology provides a general platform for proteome-scale analysis of protein turnover under various physiological and disease conditions.

A complete understanding of biological networks requires knowledge of all aspects of regulation, from gene transcription, RNA processing, translation, and localization to protein modification and turnover. However, the chemical heterogeneity of proteins, the large dynamic range of their abundance, and the absence of specific recognition reagents have hindered high-throughput systematic approaches for the analysis of protein regulation (1, 2). Thus, protein stability remains an unexploited feature for most proteins (3).

The selective degradation of proteins is critical for most cellular processes, including cell cycle progression, signal transduction, and differentiation (4–6). Control of protein turnover serves as a rapid mechanism for activation or inhibition of signaling pathways when cells respond to environmental changes. Alterations in the degradation of cancer-related proteins have important roles in cellular transformation, and multiple components of the proteolysis system are directly involved in human diseases (7–9). Furthermore, many viruses have evolved strategies to hijack the proteolytic pathway of host cells for their own benefit (10). Therefore, understanding protein turnover should have a major impact, not only for the basic science of biological regulation, but also for the development of more effective strategies to cure diseases (11, 12).

Traditional methods of measuring protein stability rely on either pulse-chase metabolic labeling or administration of protein synthesis inhibitors, followed by biochemical analysis of the abundance of the protein of interest at multiple time points during the chase period. When applying half-life analysis to a global population of proteins under a broad range of physiological or disease states, these assays are impractical. One method to increase throughput is to combine pulse-chase analysis with mass spectrometry (3). However, mass spectrometry is limited by the inability to detect low-abundance proteins, and highly regulated proteins tend to be present in low amounts. Nor can these methods be used for real-time monitoring of protein turnover in living cells with single-cell resolution, a feature important for systems level understanding of protein function (13–15). Therefore, we developed a high-throughput approach for proteome-scale protein-turnover analysis in mammalian cells, called global protein stability (GPS) analysis and described here, that overcomes many of these deficiencies (16).

Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: selledge@genetics.med.harvard.edu
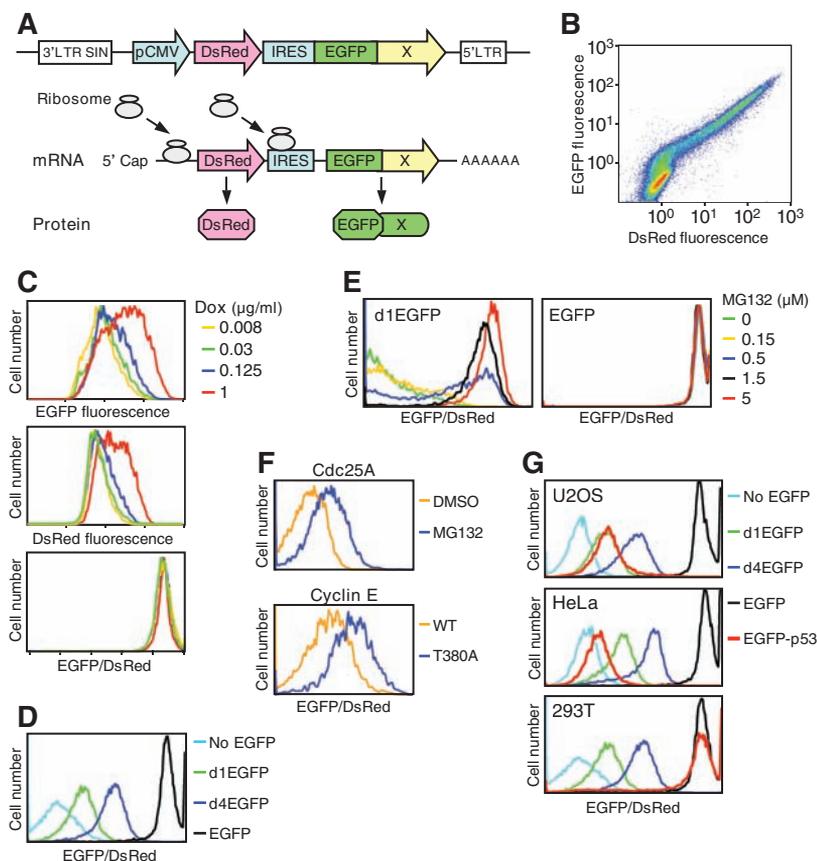
Fig. 1. Determination of protein stability by the GPS system. (**A**) A schematic representation of the reporter construct. Ribosomes can dock at the 5′ end of the bicistronic mRNA to translate DsRed or at the IRES to translate EGFP-X. (**B**) HEK 293T cells stably expressing the pCMV-DsRed-IRES-EGFP reporter were analyzed by FACS. The cluster in the lower left corner of the plot represents uninfected cells. (**C**) HeLa tetracycline-inducible (tet-on) cells with the pTRE-DsRed-IRES-EGFP expression cassette were treated with various doses of Dox and analyzed by FACS. (**D**) HEK 293T cells carrying the pCMV-DsRed-IRES-EGFP reporter cassette with EGFP, d4EGFP, or d1EGFP or without EGFP were analyzed by FACS. (**E**) HEK 293T cells expressing d1EGFP or EGFP from the DsRed-IRES-EGFP reporter cassette were treated with various concentrations of MG132 and analyzed. (**F**) (Top) HEK 293T cells expressing the DsRed-IRES-EGFP-Cdc25A reporter with or without MG132 treatment (2 μM) were compared. (Bottom) HEK 293T cells with the reporter carrying EGFP-fused wild-type or T380A mutant cyclin E were analyzed. (**G**) Overlay of the EGFP/DsRed ratios of EGFP-p53 and various EGFP-degron fusions in the three different cell lines, as indicated in the graphs.

A fluorescence-based system to monitor protein stability at the single-cell level. We established a cell-based system for measuring GPS. We built a retroviral reporter construct in which the expression cassette contains a single promoter that, with an internal ribosome entry site (IRES) (17, 18), permits the translation of two fluorescent proteins from one mRNA transcript (Fig. 1A). The first fluorescent protein, *Discosoma* sp. red fluorescent protein (DsRed), served as an internal control, whereas the second, enhanced green fluorescence protein (EGFP), was expressed as a fusion with the protein of interest (X). When integrated into the genome of cells, DsRed and EGFP-X proteins should be produced at a constant ratio because they are derived from the same mRNA (Fig. 1A). The EGFP/DsRed ratio of cells represents the stability of protein X in this system and can be quantified by fluorescence-activated cell sorting (FACS). Events that selectively affect the protein stability of EGFP-X (for example, depletion or overexpression of proteins that regulate X turnover) are expected to change the abundance of EGFP-X, but not DsRed, and thus lead to an alteration of the EGFP/DsRed ratio.

The GPS system has several important features. Because the EGFP and DsRed proteins are derived from the same mRNA, the EGFP/DsRed ratio is not affected by transcriptional regulation. In addition, the reporter construct is integrated into the genome as a single copy, and EGFP-X is expressed downstream of an IRES under the control of a ubiquitously active promoter. Because translation of the gene downstream of an IRES is less efficient than that of the upstream gene (17), this design limits overproduction of the EGFP-X protein and potential phenotypic perturbation in cells. Using a common promoter to drive reporter protein expression also reduces the range of protein abundance, a common problem for global proteomic assays. In contrast to traditional protein half-life assays that measure only the mean value of protein half-lives from a population of cells,

Fig. 2. Global protein stability signature by using microarrays. (**A**) Schematic diagram of the four steps of the experimental procedure. Reporter cells with different EGFP/DsRed ratios are shown in different colors. (**B**) The EGFP/DsRed ratio of the reporter cell library. The library cells were FACS-divided into seven sublibraries (R1 to R7) with ascending EGFP/DsRed ratio. (**C**) The EGFP-fused ORFs were PCR-amplified from the genomic DNA. RNA was transcribed in vitro from the PCR products, labeled, and used for competitive microarray hybridization. (**D**) Seven hybridizations were performed with ORFs from the total library labeled with Cy5 and ORFs from each FACS-isolated subpopulation labeled with Cy3. The EGFP/DsRed ratio of cells expressing EGFP fused to a protein of interest can be tracked by combining the hybridization results (Cy3/Cy5) from the seven microarrays. Two EGFP fusion proteins, EGFP-X (short half-life) and EGFP-Y (long half-life) are shown as examples. (**E**) Representative data. (**F**) Results for individual ORFs from microarrays [(top) percent of cells in each fraction] and FACS analysis [(bottom) EGFP/DsRed ratio] were compared.

Fig. 3. Screen validation. (**A**) Unbiased validation of microarray performance. The expected stability is derived from the EGFP/DsRed ratio from individual FACS analysis; the observed stability is the PSI derived from microarray results. The expected and observed stabilities were compared for 96 ORFs from ORFeome plate 11001. (**B**) The steady-state protein levels of 75 randomly chosen ORF-HA fusions were analyzed by Western blotting and quantified. Additional information and data related to the tested ORFs can be found in table S5. The number of ORFs with a particular range of expression was counted and plotted on a graph. S, M, L, and X represent proteins with short, medium, long, and extra-long half-lives, according to the PSI. (**C**) HEK 293T cells expressing short–half-life ORF-HA fusions were either left untreated (lane 1), treated with 2 μM MG132 for 2 hours (lane 2), or treated with MG132 for 2 hours and released for 6 hours (lane 3), and analyzed by Western blot. (**D**) Cycloheximide (CHX)-chase analysis of cells expressing medium, long, and extra-long–half-life ORF-HA proteins. Protein samples were normalized based on cell number.

the GPS system allows real-time protein stability detection at the level of individual living cells. Last, this fluorescence-based system has the potential for automation and is amenable to high-throughput proteome-scale analysis.

We integrated the GPS reporter that expresses DsRed and EGFP under the control of the cytomegalovirus (CMV) promoter and the IRES from encephalomyocarditis virus into the genome of human embryonic kidney (HEK) 293T cells and analyzed cells by FACS. Depending on the integration site, amounts of the fluorescent proteins might differ in particular cells, but the EGFP/DsRed ratio should be constant. As expected, the EGFP and DsRed signals displayed a constant ratio despite the wide range of fluorescence intensities for each protein in different cells (Fig. 1B), which suggests that this ratio is independent of transcription. To confirm this, we replaced the CMV promoter with the tetracycline response element (TRE) promoter that is controllable by doxycycline (Dox). The EGFP/DsRed ratio did not vary, despite increases in EGFP and DsRed abundance after the addition of Dox (Fig. 1C). To test whether the EGFP/DsRed ratio can be used to distinguish proteins with different half-lives, we created cell lines that express DsRed-IRES-EGFP fusions with mutants of the ornithine decarboxylase (ODC) degron that confer different half-lives (d1EGFP, $t_{1/2}$ = 1 hour; d4EGFP, $t_{1/2}$ = 4 hours; EGFP, $t_{1/2}$ = 24 hours). Cells stably expressing these fusions displayed distinct EGFP/DsRed ratios that reflect the stability of the EGFP fusions (Fig. 1D). We obtained similar results in all cell types tested (Fig. 1G). These data indicate that the EGFP/DsRed ratio is a reliable readout for protein stability in mammalian cells.

Comparison of the EGFP/DsRed ratio to other measurements of protein stability. We examined the correlation between our protein stability measurements and those reported in the literature. The ODC degron destabilizes EGFP by targeting it for hydrolysis by the proteasome (19). Consistent with that, the EGFP/DsRed ratio of cells expressing EGFP to which the degron was fused increased in response to treatment with the proteasome inhibitor MG132 in a dose-dependent manner (Fig. 1E). We also made reporter cells expressing EGFP-Cdc25A, EGFP–cyclin E and EGFP-p53 fusions. The stability of Cdc25A increased in cells treated with MG132, and cyclin E harboring the T380A mutation (substitution of Ala for Thr[380]) displayed greater stability than the wild-type protein (20) (Fig. 1F). These results are consistent with the previous discoveries that Cdc25A is a proteasome substrate and that mutation of Thr[380] in cyclin E results in defective turnover (21–24).

We measured the half-life of p53 in three different cell lines with distinct p53 turnover pathways. Comparison of the EGFP/DsRed ratio of EGFP-p53–expressing cells with that of the EGFP-degron series indicated that p53 displays distinct stabilities in these three cell lines (Fig. 1G). In human osteosarcoma U2OS cells, the half-life

of EGFP-p53 approaches that of d1EGFP. HeLa cells, which express papillomavirus E6, promote more rapid p53 degradation through E6-associated protein (E6-AP) (25) and had a lower EGFP/DsRed ratio. In contrast, EGFP-p53 is highly stable in HEK 293T cells, which express the T antigen that binds and sequesters p53 (26). In conclusion, the EGFP/DsRed ratio serves as an accurate measurement of the relative half-lives of proteins and is responsive to genetic changes.

Multiplex GPS profiling by using DNA microarray deconvolution. To apply GPS on a global scale, we designed a multiplexing strategy that integrates the power of fluorescence-based protein stability analysis with that of DNA microarray technology to rapidly obtain protein stability profiles in human cells. The four elements of this approach are shown in Fig. 2A. To create a reporter cell library, we generated a pCMV-DsRed-IRES-EGFP-fusion cDNA library in a retroviral vector using the hORFeome v1.1 library, which consists of ~8000 unique, full-length human protein–encoding open reading frames (ORFs) in

a Gateway entry vector that allows in-frame transfer of the ORFs to any expression plasmid by recombinational cloning (27). The hORFeome library gene set is an arrayed set of individual ORFs in 96-well plates, which facilitates future screen validation processes. We used the retroviral DsRed-IRES-EGFP-ORF library to produce viruses and infected HEK 293T cells to make DsRed-IRES-EGFP-ORF v1.1 reporter cell collections. To ensure that each cell carried only one reporter cassette, cells were infected at low multiplicity of infection (MOI ≈ 0.05).

We then fractionated the cell library into seven subpopulations of increasing EGFP/DsRed ratios (R1 to R7) by FACS (Fig. 2B). Because sorting of cells into distinct pools was dependent on the half-lives of the EGFP-ORF proteins they express, the stability of a particular EGFP-ORF fusion could be inferred from the distribution of cells expressing that specific fusion within the seven pools. To determine the protein stability of all 8000 ORFs, the EGFP-fused ORF sequences that serve as cell identifiers were amplified by the polymerase chain
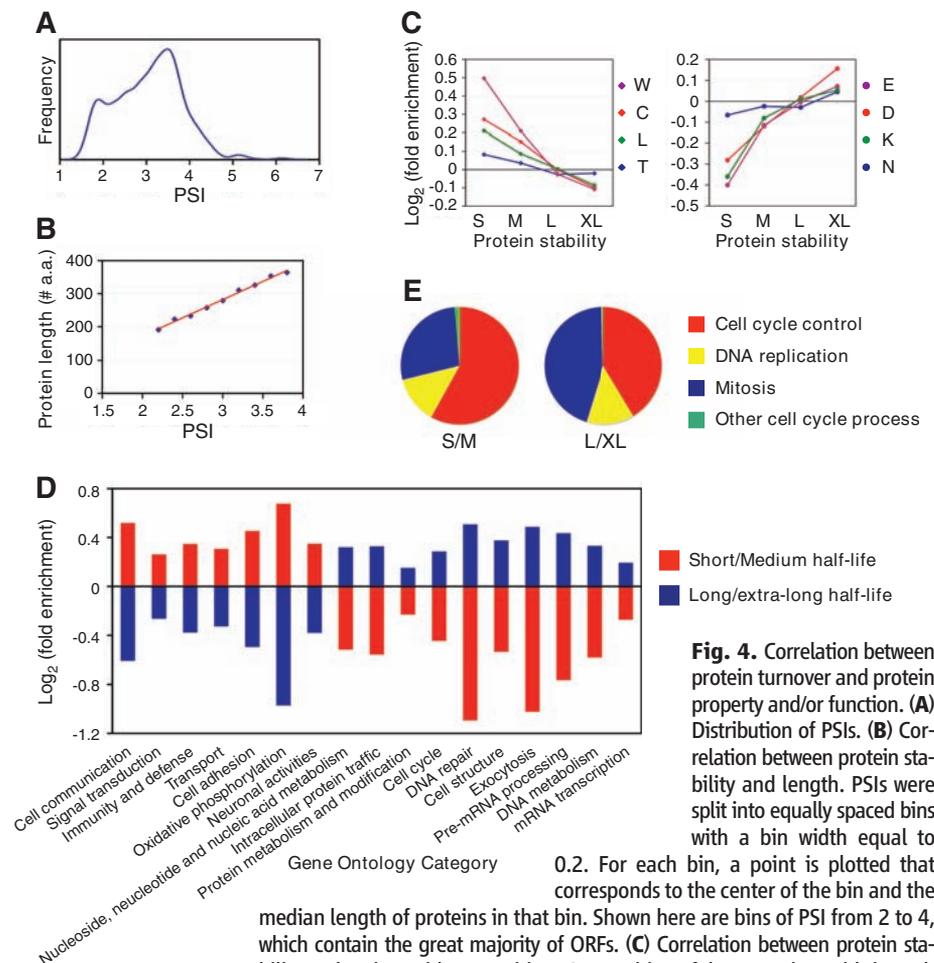
Fig. 4. Correlation between protein turnover and protein property and/or function. (**A**) Distribution of PSIs. (**B**) Correlation between protein stability and length. PSIs were split into equally spaced bins with a bin width equal to 0.2. For each bin, a point is plotted that corresponds to the center of the bin and the median length of proteins in that bin. Shown here are bins of PSI from 2 to 4, which contain the great majority of ORFs. (**C**) Correlation between protein stability and amino acid composition. Composition of the 20 amino acids in each of the four protein groups was compared with the overall composition. Shown here are four amino acids that are enriched in short–half-life proteins (left) and four that are enriched in long–half-life proteins (right). (**D**) Correlation between protein stability and function. Presented are Panther Biological Process GO categories that are inversely enriched between the short–/medium–half-life proteins and long–/extra-long–half-life proteins. (**E**) Distribution of short–/medium– (left) and long–/extra-long–half-life (right) cell cycle proteins in "cell cycle" subcategories.

reaction (PCR) from the genomic DNA of both the sorted subpopulations and the total library, and their relative abundance was quantified in microarrays (Fig. 2C). Unlike transcription profiling, which uses oligonucleotides from "mRNA" for hybridization, the ORF sequences are recovered from the "genomic DNA" of reporter cells; therefore, the microarray result reflects the composition of the library and is not affected by transcription. Because normal transcriptional microarrays are biased for probes in the 3′ untranslated region, we designed our own custom microarray to specifically detect ORF sequences (table S1).

We performed seven hybridizations of the sorted subpopulation of cells (Cy3-labeled) versus the total library (Cy5-labeled) and combined seven sets of array data to determine the abundance of different ORFs in the various pools (Fig. 2D and fig. S1A). All cells expressing a specific EGFP-ORF fusion should be distributed within the seven subpopulations after FACS, and thus, the sum of the Cy3/Cy5 ratios from seven

hybridizations for a probe should be equal to 1 (100%). To represent the protein stability information from arrays in a quantitative manner, we calculated the "protein stability index" (PSI) using the following formula: $\text{PSI} = \sum_{i=1}^{7} R_i * i$, where i is the subpopulation number (that is 1 to 7) and R_i is the fraction of the signal present for a gene in that given subpopulation i. The value of PSI ranges from 1 to 7, with a higher PSI value meaning higher relative protein stability of EGFP-fused ORFs. We also calculated the standard deviation $[\text{SD} = \sqrt{\sum_{i=1}^{7} R_i * (i - \text{PSI})^2}]$ to quantify the spread of protein stabilities in a population of cells expressing a specific EGFP-ORF fusion. The complete list of screen results is in table S2, and representative data are shown in Fig. 2E.

Screen results and validation. We carried out a series of analyses and confirmed that the overall

quality of the array hybridization is high (16) (fig. S2 and tables S3 and S4). Our data indicate that >98% of ORFs from the hORFeome v1.1 library are preserved in the reporter cell library. To examine whether the protein stability information derived from the microarray data resembled that measured by FACS analysis of individual samples, we randomly picked one plate (96 ORF clones) of the arrayed ORFeome collection for analysis. Each ORF was individually recombined into the DsRed-IRES-EGFP vector, packaged into viruses, and transduced independently into cells. The EGFP/DsRed ratio for each clone was measured by FACS and compared with the PSI calculated from the array. There was a nearly perfect correlation between the "observed stability," which reflects PSI, and the "expected stability" derived from the EGFP/DsRed ratio of individual clones, with a correlation coefficient of 0.907 (Fig. 3A). Moreover, the distribution of EGFP/DsRed ratios for a given clone derived from the array was similar to that from individual FACS (Fig. 2F). This is striking because the PSI was derived from library cells divided into only seven pools. In addition, the PSI can be affected by double integrations, cell purity after FACS fractionation, ORF preparation, or hybridization differences. In summary, these data indicate that microarray deconvolution provides a robust readout for protein stability.

One potential caveat of this technology is that the N-terminal EGFP fusion might affect protein turnover in individual cases. To address this concern, we randomly picked 75 ORFs, tagged them with a single hemagglutinin (HA) epitope at the C terminus and expressed them under the control of the elongation factor 1α (EF1α) promoter at single copy. These ORFs belong to four stability categories on the basis of our array information: short (PSI ~ 1.3), medium (~3.5), long (~5.5), and extra-long (~6.5) (table S5). We found that the abundance of ORF-HA proteins determined by Western blot analysis correlated well with the PSI of EGFP-ORF fusions measured by microarrays (Fig. 3B and fig. S3). For example, only 1 out of 19 HA-fused proteins in the short–half-life category was detected, and that amount was low. In contrast, all 18 proteins from the extra-long–half-life category were detected, and most displayed strong Western signals. The abundance of a given cellular protein is determined by the balance between its rate of synthesis and degradation, so differences in the steady-state abundance of ORF-HA fusions are likely due to differences in their turnover rates. To ensure the inability to detect ORF-HA proteins with low PSIs was indeed due to degradation, we treated cells with MG132 and observed rapid accumulation of those proteins (Fig. 3C). We also treated cells for a limited time with the protein synthesis inhibitor cycloheximide to measure the stability of medium–, long–, and extra-long–half-life proteins and found that the PSI serves as an accurate indicator of protein stability (Fig. 3D). Thus, in general, the stabilities of most proteins are not compromised by the N-terminal EGFP tag and the gateway *att* sites.

Fig. 5. Identification of proteasome substrates by comparative protein-stability profiling. (**A**) The EGFP/DsRed ratio of library cells with or without MG132 treatment. (**B**) Library cells with or without MG132 were fractionated into seven pools (R1 to R7). Seven hybridizations with untreated/Cy3 versus treated/Cy5 were performed, and data were combined. Data from cells expressing a proteasome substrate, X, are shown as an example. (**C**) Representative results. Numbers shown are log₂(Cy5/Cy3). (**D**) Representative comparison between protein stability inferred from individual EGFP/DsRed FACS measurements (top) and that from microarray data (bottom). The ORFs shown are the same as those in (C). (**E**) The steady-state amount of ORF-HA proteins from cells treated with or without MG132 was analyzed by Western blot. ORF-HA fusion proteins were expressed from the EF1α promoter.

Protein stability and protein properties/functions. We used our protein stability information to investigate whether correlations exist between particular properties of proteins and protein half-lives. To clearly separate proteins with different half-lives, we filtered out genes with a flat or bimodal stability profile and kept only 5341 ORFs that had a single sharp peak for further bioinformatics analysis (16). The global distribution of PSIs showed a bimodal pattern, with a major peak centered around 3.5 and a minor peak centered at 2 (Fig. 4A). The PSIs for d1EGFP, d4EGFP, and EGFP were 2.75, 4.19, and 6.61, respectively. With a simple linear-regression analysis based on the PSI of these three known half-life proteins, we estimated that PSIs of 2 and 3.5 approximate protein half-lives of 30 min and 2 hours, respectively. There is a significant positive correlation between protein half-life and protein length [Spearman correlation coefficient, $r = 0.28$, $P < 10^{-16}$ (Fig. 4B)], which indicated that longer proteins tend to be more stable.

To survey whether amino acid composition of proteins influences turnover, we divided ORFs into four groups according to PSI: short half-life (S: PSI < 2; 653 ORFs), medium half-life (M: $2 \leq$ PSI < 3; 1734 ORFs), long half-life (L: $3 \leq$ PSI < 4; 2442 ORFs), and extra-long half-life (XL: PSI $\geq$ 4; 512 ORFs). We observed that the amino acids W, C, L, T, F, W, Y, and V were enriched in labile proteins, whereas charged amino acids including E, D, K, N, R, and Q were enriched in stable proteins (20) (Fig. 4C). We searched for enrichment of proteins with shared Gene Ontology (GO) terms in the four half-life groups to look for correlations between protein stability and function. Many shared categories were detected between short– and medium–half-life proteins, and many categories were shared between long and extra-long–half-life proteins. By contrast, there was no overlap of enriched GO categories between short–/medium– and long–/extra-long–half-life proteins. Short–half-life proteins are enriched for membrane proteins and signal transduction proteins, whereas long–half-life proteins are enriched for cytoskeleton proteins and nuclear proteins with housekeeping functions (Fig. 4D). We also examined the relation between protein stability and cell cycle subfunctional categories, and found that the short–/medium–half-life group has a larger proportion of "cell cycle control" proteins that are generally known to be unstable; the long–half-life group has a larger fraction of "mitosis" proteins that consist of actins, tubulins, septins, and so on (Fig. 4E). Because the N-terminal EGFP tag interferes with the signal peptide for membrane anchoring, it is possible that the rapid degradation of EGFP-fused membrane proteins was caused by their mislocalization. We think that this is not the case because 14 C-terminally HA-tagged proteins with short half-lives that were tested in Fig. 3 are membrane proteins, and all were unstable regardless of the position of the tag. We also wondered whether the decreased proportion of charged amino acids

in proteins with short half-life reflected the hydrophobic nature of membrane proteins, but we obtained the same conclusion after removing all annotated membrane proteins from the data set used for the analyses.

To tightly and efficiently control the amount of proteins in cells, it is likely that transcription, translation, and protein turnover must be coordinated. Our data indicate a statistically significant positive correlation between the steady-state mRNA level and protein stability [Pearson correlation coefficient $r = 0.124$, $P < 2 \times 10^{-13}$ (fig. S4)]. A previous study in yeast also showed that low mRNA abundance correlates with instability (28).

Global identification of proteasome substrates by comparative protein stability profiling. A key goal of modern biological analyses is to determine how proteomes are remodeled in response to environmental cues. We developed a "comparative GPS profiling" strategy to identify proteins whose stabilities increased or decreased in response to stimuli. We used two pools of library cells, treated one with the stimulus of choice and sorted cells from each pool by FACS on the basis of EGFP/DsRed ratios. The relative clone distribution was deconvoluted by quantitative comparative hybridization. By comparing stability profiles from treated and untreated cells, we determined which proteins' stabilities were altered in response to the stimulus (fig. S1B).

To screen for proteasome targets, we treated library cells with MG132. EGFP itself is not degraded by the proteasome (Fig. 1E). When looking at the EGFP/DsRed profile of the entire cell library, we detected a global increase in the EGFP/DsRed ratio in response to MG132, which suggested that large numbers of ORFs encode proteasome substrates (Fig. 5A). We fractionated MG132-treated and untreated library cells into seven pools and isolated EGFP-fused ORFs for microarray hybridization. We hybridized probes derived from untreated (Cy3-labeled) versus treated (Cy5-labeled) fractions, as well as untreated versus treated total cell library, and determined the Cy5/Cy3 ratio distribution of probes for each chip (Fig. 5B). In agreement with the hypothesis that many short–half-life EGFP-ORFs might encode proteosome substrates, we observed the reduction of ORFs (Cy5 < Cy3) from cell fractions that contained proteins with short half-lives (R1 to R3) and an increase of ORFs (Cy5 > Cy3) in fractions that express proteins with higher stability (R4 to R7) after MG132 treatment. The Cy5 and Cy3 signals of most probes (95.28%) from the total library differed by less than 50%, which indicated that the overall EGFP-ORF compositions of untreated and MG132-treated cells were very similar (fig. S5). Therefore, any changes in the Cy5/Cy3 ratios represent a redistribution as opposed to overall gain or loss of signals.

We integrated and arrayed the $\log_2(Cy5/Cy3)$ ratio for each probe from seven hybridizations (R1 to R7) to search for proteasome substrates. Cells expressing an EGFP-ORF that is not a pro-

teasome target should be sorted into the same EGFP/DsRed fractions regardless of MG132 treatment, and thus the $\log_2(Cy5/Cy3)$ ratio of the probe corresponding to that ORF should be close to 0 (Cy5 $\approx$ Cy3) on all chips. In contrast, because cells expressing a specific EGFP-ORF that disappears from one cell fraction should reappear in other fractions, probes for proteasome substrates should show a negative value of $\log_2(Cy5/Cy3)$ ratio on chips representing low EGFP/DsRed ratio cells and a positive value on chips of high EGFP/DsRed ratio cells. Combining data from the seven hybridizations improved the accuracy of discriminating true-positives from false-positives caused by experimental artifacts. Representative results for proteasome substrates and nonsubstrates are shown in Fig. 5C, and the complete screen results are in table S6. Consistent with the fact that proteasome degradation is the main pathway for cellular protein turnover (29), we found that the stability of >80% of EGFP-ORF proteins increased in cells treated with MG132, and none showed a decrease.

We chose a random sample of 85 ORFs predicted to encode proteasome substrates and 6 ORFs that do not and individually measured the EGFP/DsRed ratios of DsRed-IRES-EGFP reporter cell lines expressing those ORFs in the presence or absence of MG132. In all cases examined, the EGFP/DsRed measurements were in agreement with the array results (Fig. 5D). We detected great similarities between the EGFP/DsRed profiles inferred from arrays and those obtained by FACS measurements (Fig. 5D). To demonstrate that stabilization by MG132 was independent of the EGFP fusion, we C-terminally HA-tagged 24 ORFs and determined the steady-state concentration of their encoded proteins by Western blotting in the presence or absence of MG132. The abundance increased with MG132 treatment in all cases (Fig. 5E). These data collectively support the use of comparative GPS profiling as a general and powerful platform to identify specific protein turnover events in response to environmental changes.

Protein turnover represents an underexamined dimension of proteomics. We established a novel assay as a read-out for protein half-life, GPS, and combined it with DNA microarray deconvolution to allow highly parallel multiplex analysis of protein stability of over 8000 human proteins. Protein stabilities derived from GPS profiling are consistent with those reported in the literature and respond appropriately to different genetic backgrounds. Furthermore, unbiased validations for both the global protein stability screen and the comparative protein stability profile for proteasome substrates strongly supported the validity of the GPS analysis. There are, however, limitations to GPS. In certain subsets of proteins, such as membrane proteins and mitochondrial proteins, the presence of an N-terminal EGFP tag may alter their localization and hence their turnover. This can be addressed by using a C-terminal EGFP-tagged library. Because EGFP-ORF and DsRed proteins are translated through different

mechanisms, it is possible that the product of a specific ORF affects IRES function and, thus, influences the EGFP/DsRed readout in that particular cell. There will also be cases where the EGFP fusion may affect stability by affecting folding or obscuring degron sequences. Despite these limitations, this method offers a very deep window into a critical aspect of cellular physiology. The only other comprehensive global analysis of protein turnover that has been available was performed in budding yeast by using >3800 individual cycloheximide-chase analyses (28), a method that is impractical for mammalian cells.

A number of general findings emerged from this analysis. We found a bimodal distribution of protein half-lives centered around 0.5 and 2 hours. A similar distribution was previously observed in yeast, although with a shorter scale (28), which may be explained by the shorter cell cycle of yeast (~2 hours) compared with mammalian cells (~20 hours). We also find that longer proteins are relatively more stable. One possible explanation is that cells require more resources to synthesize longer proteins and tend to protect their investment. Although PEST sequences (polypeptide sequences enriched in proline, glutamic acid, serine, and threonine) are widely thought to be associated with short–half-life proteins (30), we found no enrichment of PEST sequences in labile proteins. Instead, unstable proteins appear to be rich in amino acids that can be phosphorylated, such as tyrosine and threonine. Indeed, phosphorylation is frequently a signal for regulated protein degradation (31). Thus, we conclude that the PEST hypothesis is incorrect in a general sense.

Proteins with unstructured regions (UPRs) are susceptible to degradation by the 20S proteasome in vitro (32). However, we found no correlation between the presence of UPRs (in both length and number) and protein instability. Because many UPRs function in molecular recognition, it is possible that in vivo UPRs are no longer "unstructured" and are protected by binding to their biological targets (33). Similar observations were made in yeast (34).

This GPS technology has a number of applications. It could be used to identify mutations that affect basal protein stability, which would reveal degron or stabilization sequences. GPS profiling could also be used to identify proteins whose stabilities change in response to stimuli, as well as during developmental transitions. GPS can be used to discover ubiquitin ligases or other proteins that regulate the stability of a protein of interest by coupling GPS with loss-of-function (from RNA interference) or gain-of-function screens that alter the DsRed/EGFP ratio. Conversely, this method could be used to identify substrates of ubiquitin ligases, currently a very labor-intensive endeavor with few general solutions, as we have done with the Skp1–cullin–F-box (SCF) ubiquitin ligase (35). GPS could be coupled with chemical screens to search for compounds that destabilize a protein of

interest as opposed to inhibiting its activity by direct binding. GPS profiling could also be used to generate disease-specific protein stability signatures that may be useful for both diagnosis and elucidation of disease mechanisms. Finally, the integration of global protein stability information with other data sets will provide a global vision of regulatory networks with greater clarity and will help identify cross-talk between protein turnover and other levels of biological regulation (36, 37). Thus, GPS has opened many avenues for protein-turnover studies.

References and Notes

1. S. Ghaemmaghami et al., Nature 425, 737 (2003).
2. S. F. Kingsmore, Nat. Rev. Drug Discov. 5, 310 (2006).
3. J. M. Pratt et al., Mol. Cell. Proteomics 1, 579 (2002).
4. R. W. King, R. J. Deshaies, J. M. Peters, M. W. Kirschner, Science 274, 1652 (1996).
5. V. Jesenberger, S. Jentsch, Nat. Rev. Mol. Cell Biol. 3, 112 (2002).
6. A. Hershko, A. Ciechanover, Annu. Rev. Biochem. 67, 425 (1998).
7. K. I. Nakayama, K. Nakayama, Nat. Rev. Cancer 6, 369 (2006).
8. A. Ciechanover, A. L. Schwartz, Biochim. Biophys. Acta 1695, 3 (2004).
9. A. L. Schwartz, A. Ciechanover, Annu. Rev. Med. 50, 57 (1999).
10. L. Banks, D. Pim, M. Thomas, Trends Biochem. Sci. 28, 452 (2003).
11. G. Nalepa, M. Rolfe, J. W. Harper, Nat. Rev. Drug Discov. 5, 596 (2006).
12. Y. Sun, Cancer Biol. Ther. 2, 623 (2003).
13. J. R. Newman et al., Nature 441, 840 (2006).
14. A. Sigal et al., Nature 444, 643 (2006).
15. A. Sigal et al., Nat. Methods 3, 525 (2006).
16. Materials and methods are available as supporting material on Science Online.
17. S. M. Ngoi, A. C. Chien, C. G. Lee, Curr. Gene Ther. 4, 15 (2004).
18. T. V. Pestova et al., Proc. Natl. Acad. Sci. U.S.A. 98, 7029 (2001).
19. S. Matsuzawa, M. Cuddy, T. Fukushima, J. C. Reed, Proc. Natl. Acad. Sci. U.S.A. 102, 14982 (2005).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. M. Welcker et al., Mol. Cell 12, 381 (2003).
22. X. Ye et al., J. Biol. Chem. 279, 50110 (2004).
23. J. Jin et al., Genes Dev. 17, 3062 (2003).
24. D. M. Koepp et al., Science 294, 173 (2001).
25. M. Scheffner, J. M. Huibregtse, R. D. Vierstra, P. M. Howley, Cell 75, 495 (1993).
26. D. Lane, E. Harlow, Nature 298, 517 (1982).
27. J. F. Rual et al., Genome Res. 14, 2128 (2004).
28. A. Belle, A. Tanay, L. Bitincka, R. Shamir, E. K. O'Shea, Proc. Natl. Acad. Sci. U.S.A. 103, 13004 (2006).
29. K. L. Rock et al., Cell 78, 761 (1994).
30. M. Rechsteiner, S. W. Rogers, Trends Biochem. Sci. 21, 267 (1996).
31. T. Hunter, Mol. Cell 28, 730 (2007).
32. S. Prakash, L. Tian, K. S. Ratliff, R. E. Lehotzky, A. Matouschek, Nat. Struct. Mol. Biol. 11, 830 (2004).
33. G. Asher, N. Reuven, Y. Shaul, Bioessays 28, 844 (2006).
34. P. Tompa, J. Prilusky, I. Silman, J. L. Sussman, Proteins 71, 903 (2008).
35. H.-C. S. Yen, S. J. Elledge, Science 322, 923 (2008).
36. T. Ideker et al., Science 292, 929 (2001).
37. T. J. Griffin et al., Mol. Cell. Proteomics 1, 323 (2002).
38. We thank M. Vidal, J.W. Harper, G. Hu, and J. Jin for the hORFome v1.1 library; J. Daly and S. Lazo-Kallanian for FACS assistance; and J. Love, S. Gupta, and K. Hurov for helpful discussions. H.-C.S.Y. is a Jane Coffin Childs Memorial Fund fellow. D.M.C. is supported by the National Science Foundation. This work is supported by an NIH grant AG11085 to J.W.H. and S.J.E. S.J.E. is an investigator with the Howard Hughes Medical Institute.

Identification of SCF Ubiquitin Ligase Substrates by Global Protein Stability Profiling

Hsueh-Chi Sherry Yen and Stephen J. Elledge*

Ubiquitin-mediated proteolysis regulates all aspects of cellular function, and defects in this process are associated with human diseases. The limited number of identified ubiquitin ligase–substrate pairs is a major bottleneck in the ubiquitin field. We established and applied genetic technologies that combine global protein stability (GPS) profiling and genetic perturbation of E3 activity to screen for substrates of the Skp1–cullin–F-box (SCF) ubiquitin ligase in mammalian cells. Among the >350 potential substrates identified, we found most known SCF targets and many previously unknown substrates involved in cell cycle, apoptosis, and signaling pathways. Exploring cell cycle–stage stability, we found that several substrates used the SCF and other E3s in different cell cycle stages. Our results demonstrate the potential of these technologies as general platforms for the global discovery of E3-substrate regulatory networks.

The selective degradation of proteins is an important means of regulating gene expression and has pivotal roles in the control of various cellular processes. Ubiquitin (Ub)–mediated proteolysis is the major nonlysosomal proteolytic pathway in the cell and is required for the degradation of key regulatory proteins that include tumor suppressors, oncoproteins, and cell cycle regulators. Most proteins degraded by this pathway are first tagged with polyubiquitin chains by