PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Structural disorder serves as a weak signal for intracellular protein degradation

P. Tompa,[1] J. Prilusky,[2] I. Silman,[3] and J. L. Sussman[4]*

[1] Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary
[2] Biological Services, Weizmann Institute of Science, Rehovot, Israel
[3] Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
[4] Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT

Targeted turnover of proteins is a key element in the regulation of practically all basic cellular processes. The underlying physicochemical and/or sequential signals, however, are not fully understood. This issue is particularly pertinent in light of the recent recognition that intrinsically unstructured/disordered proteins, common in eukaryotic cells, are extremely susceptible to proteolytic degradation in vitro. The in vivo half-lives of proteins were determined recently in a high-throughput study encompassing the entire yeast proteome; here we examine whether these half-lives correlate with the presence of classical degradation motifs (PEST region, destruction-box, KEN-box, or the N-terminal residue) or with various physicochemical characteristics, such as the size of the protein, the degree of structural disorder, or the presence of low-complexity regions. Our principal finding is that, in general, the half-life of a protein does not depend on the presence of degradation signals within its sequence, even of ubiquitination sites, but correlates mainly with the length of its polypeptide chain and with various measures of structural disorder. Two distinct modes of involvement of disorder in degradation are proposed. Susceptibility to degradation of longer proteins, containing larger numbers of residues in conformational disorder, suggests an extensive function, whereby the effect of disorder can be ascribed to its mere physical presence. However, after normalization for protein length, the only signal that correlates with half-life is disorder, which indicates that it also acts in an intensive manner, that is, as a specific signal, perhaps in conjunction with the recognition of classical degradation motifs. The significance of correlation is rather low; thus protein degradation is not determined by a single characteristic, but is a multi-factorial process that shows large protein-to-protein variations. Protein disorder, nevertheless, plays a key signalling role in many cases.

Proteins 2008; 71:903–909.

Key words: intrinsically unstructured protein; natively unfolded protein; protein turnover; degradation signal; protein half-life.

INTRODUCTION

Regulated turnover is a basic device of the cell to control the levels and activities of its proteins. The underlying factors have been studied extensively, resulting in the identification of distinct signals for protein degradation. Local unfolding and exposure of hydrophobic residues is recognised by ubiquitin ligases, and the ensuing ubiquitination marks proteins for turnover by the proteasome.[1,2] Short sequence motifs, such as the destruction-box and the KEN-box, signal for a more specific degradation mechanism, that serves primarily to regulate the function of proteins involved in the cell cycle.[3,4] The N-end rule states that the half-life of a protein is determined by the nature of its N-terminal residue.[5] In an earlier attempt to establish what determines that a given protein will have a short half-life within the cell,[6,7] it was proposed that proteins containing PEST regions, that is, sequences of 10–50 amino acid residues, enriched in Pro, Glu, Ser, and Thr, tend to turn over rapidly.[8] It was thus suggested, and is currently generally accepted, that PEST regions serve as degradation signals, although the underlying mechanism(s) remained elusive.

The recent recognition that many full-length proteins and protein domains exist and function without a well-defined 3D structure, that is, that they are intrinsically unstructured/disordered (IUPs/IDPs),[9–12] may shed new light on these earlier observations. Disordered proteins are more widespread in eukaryotic proteomes than in those of eubacteria or archaea,[13,14]

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

Grant sponsor: Hungarian Research Fund; Grant number: OTKA K60694; Grant sponsor: Hungarian Ministry of Health; Grant number: ETT 245/2006; Grant sponsor: Wellcome Trust; Grant number: ISRF 067595; Grant sponsor: European Commission Sixth Framework Research and Technological Development Programme; Grant number: 031220; Grant sponsors: Israel Science Foundation; Autism Speaks; Benoziyo Center for Neuroscience; Minerva Foundation; Nalvyco Foundation; Divadol Foundation; Newman Foundation; Bruce Rosen Foundation; Kimmelman Center; Israel Ministry of Science, Culture, and Sport.
*Correspondence to: J. L. Sussman, Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: joel.sussman@weizmann.ac.il
Received 3 April 2007; Revised 25 July 2007; Accepted 10 August 2007
Published online 14 November 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21773

and carry out important functions in signal transduction and transcription regulation, often incompatible with a well-defined, stable 3D fold.[15,16] Besides numerous functional advantages, such regions are also known for their extreme proteolytic susceptibility,[9] and thus could directly signal for rapid degradation. As inferred from a range of observations, protein disorder may be mechanistically linked to protein degradation in a variety of ways. For example, it has been shown that the sites of ubiquitination in securin and cyclin B are intrinsically disordered.[4] The possible generality of this relationship is shown by the fact that short recognition motifs, such as the destruction-box, tend to fall within locally disordered regions,[17] and PEST sequences themselves correlate with protein disorder.[9,18] In addition, proteasomes can degrade disordered proteins directly, without prior ubiquitination,[19–21] and an unstructured initiation site is required even for degradation of ubiquitinated proteins.[22] This latter finding has been corroborated by the observed endoproteolytic activity of the proteasome.[19]

Thus, it is tempting to consider that structural disorder may be the missing causative factor in either spontaneous or targeted degradation *in vivo*, with or without the involvement of specific signals and/or PEST regions. This possibility has never been generally tested due to the lack of all-inclusive data on the turnover of the proteins of an entire proteome. In a recent study, however, the half-lives of proteins encompassing a very large fraction of the yeast proteome (3750 proteins, 65.2% of 5749, compare *Saccharomyces* genome database SGD[23]) were determined in a high-throughput study that utilized Western-blotting of TAP-tagged proteins.[24] This large set of data creates a unique opportunity to assess the generality of earlier scattered observations. In this study, we have systematically analysed these data to seek clues as to what determines the half-lives of proteins. Of the various possible physical factors and sequence signals tested, we found that protein disorder plays the major, albeit far from deterministic, role. Neither the presence of PEST regions, destruction-boxes and KEN-boxes nor the identity of the N-terminal residue show more than a slight correlation with half-life, whereas other structure-related factors, such as helix and coil content are reasonable predictors. An outlier group of proteins with very short half-lives, comprising 4.5% of the total, do, however, possess an increased PEST motif content. It should be noted that in all cases the significance of correlation is rather low; thus protein degradation is not determined by a single characteristic, but is a multi-factorial process that shows large protein-to-protein variations. Protein disorder, nevertheless, plays a key signalling role in many cases.

DATA AND PREDICTIONS

Data

Values for protein half-lives were taken from the database generated in the high-throughput study of all yeast gene products by Western-blotting of TAP-tagged genes.[24] The data are available as supporting information on the PNAS web site www.pnas.org, and were used with two modifications. First, proteins with a half-life of exactly 300 min were discarded, because this value was assigned in the original publication to proteins for which the degradation curve could not be fitted by an exponential decay function, or for which fitting yielded a negative half-life (~350 proteins). We also removed six outliers, with extremely long half-lives (>6000 min), because, of necessity, they have a very high error range due to their being derived from measurements of protein abundances 0, 15, and 45 min after cycloheximide treatment of the cells,[24] and also because their inclusion would disproportionately bias the statistics for proteins with much shorter half-lives. This procedure results in 3750 proteins, which represents 65.2% of the 5749 proteins in the yeast proteome.[23] Thus, about 2000 proteins are missing from our analysis, either because they had expression levels below the detection limit of the Western blot,[24] or because the tag had critically interfered with their expression/folding, thus precluding normal behaviour. Accordingly, our results apply to all proteins only to the extent this dataset is representative of the whole yeast proteome.

Copy numbers

Copy numbers of the major intracellular proteases/proteolytic systems were taken from the high-throughput studies of exponentially growing yeast cells,[25] where proteins were TAP-tagged, and their copy numbers determined by quantitative Western-blotting. First, proteases were collected from the *Saccharomyces* genome database SGD[23] available at http://www.yeastgenome.org/, by searching for proteins with the GO terms, Biological process—"protein catabolic process" and Molecular function—"peptidase activity." Hits obtained with this search are listed in Table SI (Supplementary material). Systematic SGD ORF names were then used to extract cellular copy numbers from Supplementary Table SII of Ref. 25.

Predictions

Various physical and sequential features of the proteins were calculated or predicted as follows. Intrinsic structural disorder was predicted by two algorithms, IUPred,[26,27] available at the IUPred server at http://iupred.enzim.hu/, and FoldIndex,[28] available at http://bip.weizmann.ac.il/fldbin/findex. Both predictors were used to calculate the number of disordered residues (NumDisRes), the number of disordered segments longer than 30 consecutive residues (NumSeg), and the disorder score averaged for the entire protein (Score). Secondary structure elements, that is, helix, sheet, turn, and coil, were computed by an approximation using the GOR method as implemented in the program "garnier" from

the EMBOSS-FASTA package.[29] PEST motifs were calculated by the program "epestfind," also from the EMBOSS package. Destruction-box content was estimated by the "Destruction Box Motif (D box) Finder" algorithm by Dana Reichmann, Shmuel Pietrokovski and Michael Brandies, available at http://bioinfo2.weizmann.ac.il/~danag/d-box/main.html, whereas KEN box content was determined by a sequence search for the KENxxxN/D motif. Low-complexity regions (LowComplex) were computed with program "seg" from the NCBI,[30] and protein extinction coefficients (pExtCoeff) were calculated as described in.[31] The influence of the N-terminal residue (N-end) was estimated as suggested in.[5] The basic idea is that newly synthesized proteins harbour a stabilizing N-terminal Met, and do not function as N-end substrates. Methionine aminopeptidase may cleave off Met and generate a novel N-terminus, which may be either stabilizing or destabilizing. The possible magnitude of its effect on the *in vivo* half-life can be assessed by the ProParam tool at the ExPaSy server, http://www.expasy.ch/tools/protparam-doc.html.

RESULTS

The correlation of half-lives with various physical and sequential features was studied in three different ways.

First, quantitative measures of physical and sequential features of proteins (e.g., number of residues, cumulative length or frequency, cf. Data and predictions) were plotted against the logarithms of their half-lives, and possible correlations were assessed by linear fitting. We plotted data against the logarithms, as opposed to the actual values, of half-lives for two reasons. First, on a linear scale data are spread out by almost four orders of magnitude, and any fit would have been dominated by a few proteins with very long half-lives, effectively leaving short half-life proteins out of consideration. The other reason is that, by using a log scale, we can explore correlations with a parameter that may be considered a pseudo activation-energy of destruction because of the Arrhenius relation. Of course, this approximation assumes that the destruction process is governed by a single transition state, which is a vast simplification. The results of fits are shown for molecular weight (Fig. 1A), number of residues in low-complexity regions (Fig. 1B), for protein disorder characterized by the number of long (>30 consecutive residues) disordered segments predicted by the FoldIndex algorithm (Fig. 1C), and for several other features (Fig. S1, Supplementary material). Because the ranges of the actual measures vary widely (e.g., the disorder score covers a range of 0–1, whereas M_w varies between 5000 and 360,000) the resulting slopes are not directly comparable. For purposes of comparison, all slopes have been normalized to the respective range of features, and, together with the r^2 values characteristic of the statistical significance of the correlations, are presented in Table SII (Supplementary material).

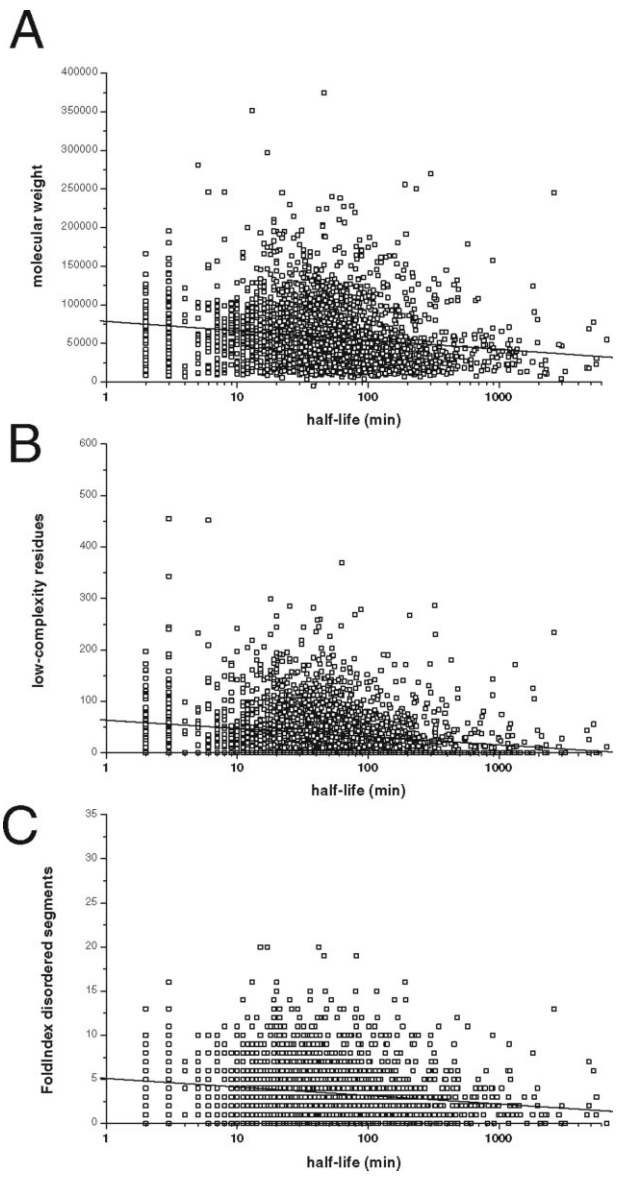

Figure 1

*Correlation of physical and sequence features with protein half-lives. The molecular weights (**A**), number of residues in low-complexity regions (**B**), and number of long disordered segments (**C**) of yeast proteins are plotted vs. the logarithms of their half-lives.[24] To determine the correlation of the features with the half-lives, linear plots were fitted to the data. Similar plots for the other physical and sequence features are presented in Fig. S1 (Supplementary material), and the resulting parameters of the fits are shown in Tables I and SII. For details of the calculations compare Data and predictions.*

These values provide important insights into what factors may determine the physiological half-lives of proteins. For most of the features, the slope is very close to zero, with the possible exception of M_w and of features characterizing disorder, either determined as the overall percentage of residues predicted to be disordered, or as

Table I
Correlation with Protein Half-Lives of Physical and Sequence Features Normalized to Protein Length

Feature	Slope	Intercept	σ	r^2
IUPredNumDisRes	−0.057415	0.262528	0.214129	0.018882
FindexNumDisRes	−0.043058	0.432077	0.254107	0.007628
Helix	0.041956	0.337629	0.131130	0.026674
Coil	−0.033603	0.430604	0.098898	0.029979
AA_V	0.033521	0.305959	0.110802	0.023915
AA_N	−0.031084	0.307256	0.091252	0.030127
AA_A	0.024280	0.194770	0.091929	0.018331
Turn	−0.023188	0.295853	0.084685	0.019676
IUPredNumSeg	−0.021615	0.140695	0.116764	0.009090
AA_S	−0.021021	0.210991	0.057387	0.034674
AA_K	0.018260	0.249729	0.094822	0.009829
DBox	0.017625	0.023693	0.075272	0.014465
Lowcomplex	−0.016040	0.101216	0.088575	0.008703
Avgreswgt	0.015597	0.030527	0.050930	0.024490
AA_G	0.015427	0.184865	0.083405	0.009076
AA_P	−0.013970	0.268134	0.092579	0.006059
AA_H	−0.013450	0.291395	0.124712	0.003104
AA_E	0.013225	0.254526	0.098336	0.004819
Destruction_Box	0.013068	0.029891	0.081720	0.006799
IP	0.011570	0.019731	0.043652	0.018459
AA_T	−0.011159	0.226788	0.067741	0.007212
FindexNumSeg	−0.010891	0.196081	0.107928	0.002719
PEST	−0.010614	0.060665	0.082091	0.004455
AA_I	0.006965	0.375732	0.108739	0.001097
AA_L	0.005526	0.414544	0.103686	0.000760
AA_F	−0.005323	0.266431	0.100114	0.000756
AA_D	−0.004424	0.373386	0.116694	0.000385
IUPredScore	0.004007	0.015918	0.029587	0.004886
AA_Y	−0.003689	0.278520	0.106539	0.000321
Sheet	−0.003477	0.421165	0.121064	0.000221
AA_Q	−0.002299	0.162265	0.074095	0.000258
AA_R	0.002085	0.111717	0.048040	0.000504
pExtCoef	−0.001121	0.249374	0.123438	0.000022
AA_KR	−0.001080	0.038542	0.040528	0.000190
AA_M	0.000798	0.134460	0.064648	0.000041
FindexScore	0.000797	0.805596	0.017641	0.000546
KEN_Box	−0.000448	0.010118	0.062057	0.000014
AA_W	0.000372	0.193496	0.143478	0.000002
N-end	0.000055	0.594136	0.487874	0.001333
AA_C	−0.000037	0.064762	0.053870	0.000000

The correlations of calculated or predicted parameters (defined in Data and predictions) with half-lives were determined as shown in Figures 1 and S1, and the resulting parameters of the linear fits are displayed in Table S2. Because of the strong dependence of half-lives on M_w, feature values are normalized to the number of amino acid residues in the protein, and linear fits are used to determine the resulting correlations with half-lives. The parameters of the fits are given here in descending order of the absolute values of the slopes.

the number of long disordered segments. The correlations with half-life of certain structural parameters, such as turns, coils and sheets, and the propensities of certain amino acids, such as P, S, and T, also deviate from mere chance occurrence. The importance of these trends is, however, difficult to assess, due to the strong influence of chain length (M_w), as a consequence of which practically all features show a negative correlation with half-life (cf. the negative values for all amino acid propensities, which should rather scatter around a mean of zero). To clarify whether any of the features affect protein turnover irrespective of chain length, that is, as a signal recognised specifically by some component(s) of the degradation machinery, the physical/sequential features were normalized with respect to chain length, and the resulting parameters of fit are presented in Table I.

It can be seen that the degree of disorder correlates inversely with half-life, and that there is a positive correlation with helical content, as well as a negative correlation with the propensity for coil formation. As to propensities associated with individual amino acids, a slight negative correlation is seen with amino acids associated with PEST motifs, and a positive correlation with some others (such as V and A). The strongest indicator related to disorder is the percentage of disordered amino acids, with less significance of the content of long disordered segments, and of the average of the disorder score. Classical degradation signals, such as the PEST region, destruction-box, KEN box, and the N-end residue are weak correlators, their presence having hardly any influence on protein turnover. In general, the level of significance of all the correlations is very low, due to the large scatter of individual scores for any given half-life value (Figs. 1 and S1), as reflected quantitatively in the low values of the correlation coefficients (Table I, r^2).

A second way of probing the possible correlation of half-lives with the physical and sequential features is to compare proteins displaying very short half-lives with the rest of the proteins in the database. It was noted[24] that for ∼96% of the proteins the half-lives show a normal distribution around a mean value of 43 min, but ∼4% display very short half-lives, within a range of 1–3 min. It is reasonable to assume that this deviation from the normal distribution is due to some unique characteristic(s) of these latter proteins. To probe for such putative characteristics, averages of the features of proteins with half-lives of 1–3 min and of 4–6000 min were calculated and compared (see Fig. 2). This comparison shows some significant deviations from the correlations found for all proteins (Table I). For the short half-life category, no correlation is seen with structural features, such as helix, sheet or turn content, and amino-acid frequencies also have no discriminative power in the short half-life group. Showing the most significant correlation with very short half-lives are the number of disordered residues and long disordered segments, and the presence of PEST regions. All three of these features are ∼2-fold more probable in the short half-life group than in other proteins, making them the strongest predictors of very short half-lives.

A third way of determining if any of the features correlate with half-lives of proteins is to compare two large groups of proteins, one optimized for maximum production and maintenance efficiency (production cluster), and the other for regulatory flexibility (regulation cluster) as defined by Belle *et al.*[24] The two clusters do not differ merely in their half-lives, but also in production rate and abundance. Practically all the features considered are much larger for the regulation cluster than for the

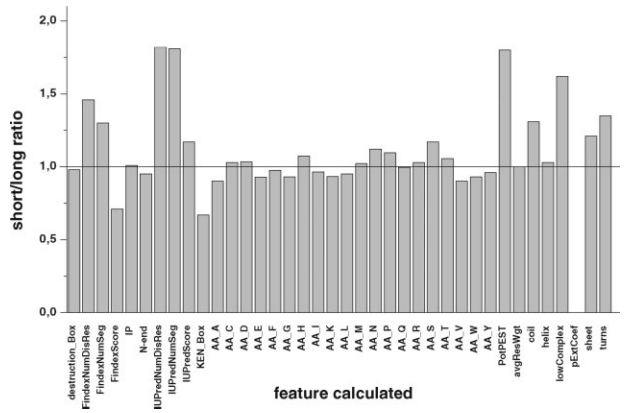

Figure 2

Differences between physical and sequence features of proteins of extremely short and of longer half-lives. The means of various parameters were determined for proteins with conspicuously short half-lives (1–3 min) and for all proteins with longer half-lives (4–6000 min). Here the differences between the means are plotted. The parameters are as defined in Data and predictions.

production cluster (data not shown), which resembles the overall negative correlation of practically all these features with half-lives (Table SII), and is a consequence of the proteins in the regulation cluster being significantly larger than those in the production cluster. Thus, to further clarify whether the frequency of any feature distinguishes between the two groups, the features were recalculated after normalizing for the lengths of the polypeptide chains (see Fig. 3). The pattern is, overall, similar to that observed prior to normalization, but some differences arise. The most strongly distinguishing features are the higher frequencies of disordered residues and of disordered segments in the regulation cluster, derived by use of the IUPred algorithm. This observation is in good correlation with prior reports that protein disorder prevails in regulatory and signalling proteins.[13–15] PEST region content no longer distinguishes between the two groups of proteins, and certain features, such as the well-established degradation signals, destruction-box and KEN-box, are of conspicuously low frequency in the regulation group.

The low correlation of physical/sequential features with half-lives in any of the above comparisons suggests that protein turnover is not determined solely by the properties of the target proteins, but at least as much by the regulation of the proteases/proteolytic systems responsible for their turnover. This issue is particularly pertinent since structural disorder renders proteins highly sensitive to degradation *in vitro*, in contrast to the *in vivo* situation. To formally demonstrate that there are hardly any nonregulated proteolytic activities in the cytoplasm, we have collected the proteases/proteolytic systems with the highest copy numbers from a recent high-throughput expression study[25] (Table SI, Table II). As can be seen, almost none of these systems lack control in the cell, their activities being regulated by various means, such as ubiquitination, localisation, special substrate requirement, or posttranslational modification. Thus, uncontrolled housekeeping proteolytic activity in the cell is probably negligible compared with the total activity of controlled proteases.

DISCUSSION

Although protein turnover is an essential component of the regulation of cellular function, its basic molecular determinants are far from well understood. The recent high-throughput characterization of half-lives of the entire yeast proteome[24] created a unique opportunity to track down what physicochemical factors and/or sequence signals predispose proteins for rapid cellular disappearance.

We found that half-life depends primarily on the length of the polypeptide chain, that is, on the number of potential cleavage sites, which suggests that protein turnover has a significant nonregulated component that stems directly from the presence of cleavable sites along the polypeptide chain. Part of this effect may also be due to the fact that longer proteins contain more residues in conformational disorder, thus presenting more susceptible sites for the intracellular degradation machinery. In this sense structural disorder represents an extensive signal that exerts its effect by its physical presence.

The fact that structural disorder also acts in an intensive, specific manner is demonstrated by the fact that dis-

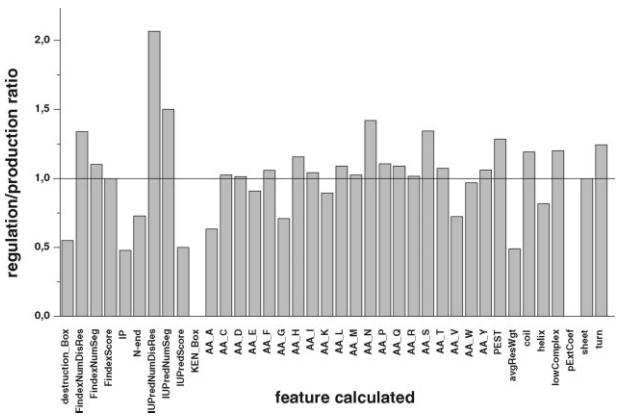

Figure 3

Differences between physical and sequence features of proteins optimized for production and regulation. The means of various parameters were determined for proteins optimized for maximum production and maintenance efficiency (production cluster) and for regulatory flexibility (regulation cluster), as defined in Ref. 24. Here, the differences between the means of the physical and sequence features are plotted. The parameters are as defined in Data and predictions.

Table II
Copy Numbers and Regulatory Mechanisms of the Major Intracellular Proteolytic Systems

SGD ORF	Protein (protease)	Mol/cell	Regulation
YMR297W	Vacuolar carboxypeptidase Y (proteinase C)	44049.76	Localisation, proenzyme activation
YER012W	20S proteasome beta-type subunit*	21778.28	Ubiquitination
YPR024W	Subunit of the mitochondrial inner membrane i-AAA protease	20138.98	Localisation, special class of unfolded substrates
YOL038W	20S proteasome alpha-type subunit*	16761.48	Ubiquitination
YFR004W	Metalloprotease subunit of the 19S	16352.99	Localisation, substrate specificity
YBL022C	Mitochondrial ATP-dependent protease	14538.98	Localisation, ATP dependence, substrate specificity
YMR089C	Component of the mitochondrial inner membrane m-AAA protease	11466.02	Localisation, special class of unfolded substrates
YCL057W	Zinc metalloendopeptidase, mitochondrial	8914.00	Localisation
YGL203C	Protease involved in the processing of killer toxin	8553.19	Membrane localisation, carboxypeptidase B-like activity
YDR430C	Lysine-specific metalloprotease of the mitochondrial	6139.31	Mitochondrial intermembrane localisation
YKL103C	Vacuolar aminopeptidase	5727.81	Localisation, proenzyme activation
YNL045W	Leucyl aminopeptidase (leukotriene A4 hydrolase)	5592.62	Aminopeptidase activity, localisation, lipid activation
YLR120C	Aspartic protease, attached to the plasma membrane	5436.17	Membrane localisation, specific inhibitor
YER078C	Metallopeptidase, mitochondrial matrix	5083.79	Localisation, peptidase activity
YBR286W	Vacuolar aminopeptidase Y	4738.82	Localisation, proenzyme activation

Copy numbers of the major intracellular proteases/proteolytic systems were taken from systematic studies of protein expression carried out by high-throughput TAP-tag quantitative Western-blotting studies (cf. Supplementary table S1). Here, the top 15 hits of the largest copy numbers are shown, alongside their mechanisms of control, as taken from references at the SGD site (http://www.yeastgenome.org/).
*Various α and β subunits of the proteasome have been neglected, with only one α and one β subunit representing the entire proteasome; otherwise, 8 proteasome subunits would have been among the top 15 proteases (cf. Table S1).

order normalized for protein length is the major feature in determining a short half-life for proteins. In other words, the presence of disorder renders proteins sensitive to degradation irrespective of the actual length of the chain. It is possible that the presence of protein disorder is mandatory for the recognition of classical short degradation signals, such as the destruction-box or the KEN-box,[4,17] but it is also possible that a disordered segment serves as a kind of conformational recognition element, which is deciphered due to its lack of a stable fold, irrespective of its actual sequential content. The observation that ubiquitinated proteins require disordered initiation sites for degradation,[22] and that disordered proteins may be directly degraded by the 20S proteasome without prior ubiquitination,[19,20] is in line with this notion. Although the correlation with disorder is rather weak, the importance of this observation is underlined by the fact that the presence of bona fide degradation signals, such as the destruction-box or KEN-box,[4,17] shows no correlation with the half-life at all.

It should not be overlooked that protein half-lives show only very weak dependence on structural disorder. The importance of this dependence is, nevertheless, underscored by its relationship with some structural features, in particular with helical content, that shows positive correlation, and with coil content, that shows negative correlation. Of possible further relevance is that PEST regions also show some correlation with overall half-lives, and with very short-lived proteins. PEST segments have been traditionally considered as regions involved in signalling protein degradation,[6,7] although the underlying mechanisms have remained elusive. The recent suggestion that PEST regions are correlated with, but do not fully correspond to, regions of protein disorder,[9,18] is in good agreement with this observed behaviour.

There can be several reasons why protein disorder is not truly deterministic in protein turnover *in vivo*, given the extreme proteolytic susceptibility it confers on IUPs *in vitro*. Protein degradation is regulated by post-translational modifications, the sites of which frequently occur within regions of disorder.[17,32] Because these regions are often short relative to total protein length, their disorder may be highly correlated with destruction, which may remain largely unnoticed, however, due to the lack of general knowledge as to where all these sites are located in proteins. In addition, many IUPs may be protected by complexation with other protein or nucleic acid molecule(s). IUPs indeed are known often to carry out their function by molecular recognition,[11,33] and a high level of disorder has been reported in hub proteins, that is, proteins with multiple partners in the interactome.[34–36] For these, degradation may be regulated by the disassembly of the complex, and also by activation of the protease targeted at the protein. As shown by our compilation of the intracellular proteolytic enzymes/systems, intracellular protease action is tightly controlled; thus intrinsic susceptibility to proteolysis is not necessarily rate-limiting for protein degradation *in vivo*.

Overall, we may conclude that the most significant and general feature that correlates with intracellular degradation is protein disorder. The very weak correlation observed, however, warns that degradation is a multi-factorial process to which several signals, physical features, and regulatory mechanisms contribute.

ACKNOWLEDGMENTS

J. L. Sussman is the Pickman Professor of Structural Biology.

REFERENCES

1. Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Biochem 1998;67:425–479.
2. Hershko A. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death Differ 2005;12:1191–1197.
3. Pfleger CM, Kirschner MW. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev 2000;14: 655–665.
4. Cox CJ, Dutta K, Petri ET, Hwang WC, Lin Y, Pascal SM, Basavappa R. The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. FEBS Lett 2002;527:303–308.
5. Mogk A, Schmidt R, Bukau B. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol 2007;17:165–172.
6. Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 1986; 234:364–368.
7. Rechsteiner M, Rogers SW. PEST sequences and regulation by proteolysis. Trends Biochem Sci 1996;21:267–271.
8. Dice JF. Molecular determinants of protein half-lives in eukaryotic cells. FASEB J 1987;1:349–357.
9. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.
10. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.
11. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579:3346–3354.
12. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 2005;18:343–384.
13. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.
14. Tompa P, Dosztanyi Z, Simon I. Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. J Proteome Res 2006;5:1996–2000.
15. Iakoucheva L, Brown C, Lawson J, Obradovic Z, Dunker A. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.
16. Minezaki Y, Homma K, Kinjo AR, Nishikawa K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol 2006;359:1137–1149.
17. Fuxreiter M, Tompa M, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 2007;23:950–956.
18. Singh GP, Ganapathi M, Sandhu KS, Dash D. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. Proteins 2006;62:309–315.
19. Liu CW, Corboy MJ, DeMartino GN, Thomas PJ. Endoproteolytic activity of the proteasome. Science 2003;299:408–411.
20. Asher G, Reuven N, Shaul Y. 20S proteasomes and protein degradation "by default." Bioessays 2006;28:844–849.
21. Tsvetkov P, Asher G, Paz A, Reuven N, Sussman JL, Silman I, Shaul Y. Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. Proteins, E-pub ahead of print, 10.1002/prot.21614.
22. Prakash S, Tian L, Ratliff KS, Lehotzky RL, Matouschek A. An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat Struct Mol Biol 2004;11:830–837.
23. Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Feierbach B, Fisk DG, Hirschman JE, Hong EL, Issel-Tarver L, Nash R, Sethuraman A, Starr B, Theesfeld CL, Andrada R, Binkley G, Dong Q, Lane C, Schroeder M, Botstein D, Cherry JM. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. Nucleic Acids Res 2004;32:D311–D314. Database issue.
24. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK. Quantification of protein half-lives in the budding yeast proteome. Proc Natl Acad Sci USA 2006;103:13004–13009.
25. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. Global analysis of protein expression in yeast. Nature 2003;425:737–741.
26. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and instrinsically unstructured proteins. J Mol Biol 2005;347:827–839.
27. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21:3433–3434.
28. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 2005;21:3435–3438.
29. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000;16:276–277.
30. Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem 1994;18:269–285.
31. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 1989;182: 319–326.
32. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004;32:1037–1049.
33. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.
34. Ekman D, Light S, Bjorklund AK, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? Genome Biol 2006;7:R45.
35. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2006;2:e100.
36. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. J Proteome Res 2006;5:2985–2995.