Sequence composition of disordered regions fine-tunes protein half-life

Susan Fishbain1,2, Tomonao Inobe2,3, Eitan Israeli2, Sreenivas Chavali4, Houqing Yu1,2, Grace Kago1, M Madan Babu4 & Andreas Matouschek1,2

The proteasome controls the concentrations of most proteins in eukaryotic cells. It recognizes its protein substrates through ubiquitin tags and initiates degradation at disordered regions within the substrate. Here we show that the proteasome has pronounced preferences for the amino acid sequence of the regions at which it initiates degradation. Specifically, proteins in which the initiation regions have biased amino acid compositions show longer half-lives in yeast than proteins with unbiased sequences in the regions. The relationship affects the degradation rates of proteins in vitro, can explain the unexpected stability of natural proteins in yeast and may affect the accumulation of toxic proteins in disease. We propose that the proteasome’s sequence preferences provide a second component to the degradation code and may fine-tune protein half-life in cells.

The ubiquitin-proteasome system (UPS) adjusts cellular protein concentrations by selecting specific proteins for destruction and hydrolyzing them into small peptides. The proteasome is the protease at the center of this system and is a 2.5-MDa multi-subunit particle. Proteins are targeted to the proteasome through a two-part degradation signal or degron. The degron consists of a proteasome-binding tag, typically a polyubiquitin chain, and a disordered region at which the proteasome engages its substrate and initiates degradation1–3. We call this disordered region the initiation site or region. Once the substrate is properly engaged, the proteasome sequentially degrades it4. Many proteins are ubiquitinated but not degraded by the proteasome in cells, and ubiquitin tags serve as signals in processes such as membrane trafficking and chromatin rearrangements5. In some cases the proteasome likely does not recognize the ubiquitinated protein, perhaps because the ubiquitin tag is first recognized by receptors associated with a different cellular process6. In other cases, the proteasome may recognize a protein but fail to initiate degradation, e.g., if the protein lacks a disordered region of sufficient length at the appropriate location6,7.

In this study we investigated why some proteins resist proteasomal degradation despite containing both a proteasome-binding tag and a disordered region. We first analyzed three specific proteins: the diffusible proteasome substrate receptor Rad23, the ubiquitin-conjugating enzyme Cdc34 and a fragment of huntingtin (Htt) protein that accumulates in people with Huntington’s disease (HD).1 We propose that the proteasome is unable to initiate degradation of these proteins because it does not recognize the amino acid sequences of their disordered regions. We went on to investigate the proteasome’s amino acid sequence preferences by comparing the degradation of proteins that differ only in their initiation regions. We found that degradation rates correlate with the amino acid composition of the initiation regions: the more diverse an amino acid sequence, the better it is recognized by the proteasome. The same correlation holds on a genomic scale between the half-lives of approximately 4,000 mouse proteins and the sequence of their disordered regions. We propose that the amino acid sequence composition of disordered regions fine-tunes protein half-life and that genetic mechanisms that generate diversity in sequence composition may represent a source of phenotypic variation.

RESULTS
Rad23 escapes proteasomal degradation
Rad23 and the related Dsk2 and Ddi1 proteins contain a ubiquitin-like (UbL) domain, which binds to the proteasome, and one or more ubiquitin-associated (UBA) domains, which bind to polyubiquitin chains in proteasome substrates. The substrates are degraded, whereas the UbL-UBA proteins escape intact8–10. UbL-UBA proteins contain disordered linker regions, but the linkers are flanked by folded domains, which impede degradation11 (Fig. 1a). We noticed that the amino acid sequences of the linkers in the yeast UbL-UBA proteins are strongly biased in that some amino acids occur more frequently than expected, though different amino acids dominate in different linkers8,12 (see also their sequence annotation in the protein family database Pfam8,12). Therefore, we asked how well the proteasome would initiate degradation at the linkers if its access were not constrained by flanking domains.

The three linkers in Rad23 range from 60 to 68 amino acids in size (Fig. 1a). Polypeptide tails of this length at the C termini of model substrate proteins support efficient proteasomal initiation and degradation6. We tested whether the Rad23 linkers would also support degradation of these same model substrates. The proteins

1Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas, USA. 2Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA. 3Frontier Research Core for Life Sciences, University of Toyama, Toyama, Japan. 4Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. Correspondence should be addressed to A.M. (matouschek@austin.utexas.edu).

Received 17 April 2014; accepted 19 December 2014; published online 2 February 2015; doi:10.1038/nsmb.2958
consisted of the compact 17-kDa protein *Escherichia coli* dihydrofolate reductase (DHFR), and we targeted them to the proteasome by fusing four ubiquitin moieties in series to the proteins’ N termini\(^3,13\). The proteasome did not degrade the resulting ubiquitin-DHFR proteins unless we also attached a disordered tail to the C terminus of DHFR to allow the proteasome to engage its substrates and initiate degradation\(^4,6,7\). For this study, we constructed five ubiquitin-DHFR test substrates: in three of them we attached each of the Rad23 linkers to the C terminus of DHFR (Ub\(_4\)-DHFR-L1, Ub\(_4\)-DHFR-L2 and Ub\(_4\)-DHFR-L3); in one we attached a 102-amino-acid–long tail derived from *Saccharomyces cerevisiae* cytochrome \(b_2\), which we knew would allow the proteasome to initiate degradation\(^6\) (Ub\(_4\)-DHFR-102); and in one we attached a tail, also derived from cytochrome \(b_2\), that is too short to function as a good initiation region\(^6\) (Ub\(_4\)-DHFR-15) (Fig. 1b and Supplementary Table 1). We synthesized the proteins by *in vitro* transcription and translation in *E. coli* extract, and each presented to purified yeast proteasome in the presence of ATP and an ATP-regenerating system. The proteasome degraded the proteins with the established initiation site (Ub\(_4\)-DHFR-102) but not any of the proteins with Rad23 linkers (Ub\(_4\)-DHFR-L1, Ub\(_4\)-DHFR-L2 or Ub\(_4\)-DHFR-L3) or the short tail (Ub\(_4\)-DHFR-15) as initiation regions (Fig. 1b). Thus, we concluded that some property of the Rad23 linkers prevents the proteasome from engaging the substrates and initiating degradation. This property serves as an additional mechanism protecting Rad23 from degradation.

Cdc34 lacks a good proteasomal initiation site
The amino acid composition of the Rad23 linkers is biased in the sense that some amino acids are noticeably over-represented while others are under-represented\(^6\). We wondered whether biased amino acid sequences occurred in other natural proteins that unexpectedly escape degradation by the proteasome. Cdc34 is a ubiquitin-conjugating enzyme (E2), and it can be ubiquitinated like many other UPS-related proteins\(^14,15\). The ubiquitin moieties attached to Cdc34 are linked through Lys48 of ubiquitin as found in bona fide proteasome degrons\(^16\), and Cdc34 contains a C-terminal disordered region of 130 amino acids, yet Cdc34 is not degraded. The amino acid composition of Cdc34’s disordered tail is strongly biased, with 50 of the 130 amino acids in the tail being aspartate or glutamate residues. These residues are concentrated in a 99-amino-acid–long acidic region (Fig. 2a). We asked whether the proteasome would be able to engage Cdc34 at its acidic region and, if not, whether the acidic region would act in a dominant manner and prevent the proteasome from initiating degradation at other regions within the protein.

We first tested whether the isolated C-terminal tail of Cdc34 would allow the proteasome to initiate degradation on a model substrate *in vitro*. We constructed three test substrates as described above: ubiquitin DHFR without a tail (Ub\(_4\)-DHFR), ubiquitin DHFR with a good initiation region (Ub\(_4\)-DHFR-102) and ubiquitin DHFR with the C-terminal 86 amino acids of Cdc34 (Ub\(_4\)-DHFR–acidic tail). We then presented them to yeast proteasome. As expected, the Ub\(_4\)-DHFR protein without a tail remained stable, and the proteasome effectively degraded Ub\(_4\)-DHFR-102 (refs. 3,6,7) (Fig. 2b). Ub\(_4\)-DHFR–acidic tail also remained stable (Fig. 2b), suggesting that the Cdc34 acidic region does not allow the proteasome to initiate degradation.

**Figure 1** Rad23 linker regions do not act as efficient initiation sites. (a) Sketch of *S. cerevisiae* Rad23 protein. RBD, Rad4-binding domain. L1–L3 indicate linkers connecting compact domains; their lengths in number of amino acids are shown below. (b) *In vitro* degradation kinetics for model proteins by purified *S. cerevisiae* proteasome. The proteins consist of four ubiquitin domains fused in series followed by an *E. coli* DHFR domain and different disordered tails at the C terminus. 15 and 102, 15 or 102-amino-acid–long tails derived from *S. cerevisiae* cytochrome \(b_2\). L1, L2 and L3 are the linkers defined in (a); the graph plots the amount of protein estimated by electronic autoradiography in SDS-PAGE gel bands (shown on the right) over time as a percentage of the initial protein amount as described in Online Methods. Data points represent mean values determined from three repeat experiments; error bars indicate s.e.m.

**Figure 2** The acidic region of Cdc34 does not support proteasomal degradation. (a) Schematic representation of a ubiquitinated Cdc34 protein. The amino acid sequence of the acidic region is shown with aspartates and glutamates in green, UBC, ubiquitin-conjugating. (b) *In vitro* degradation kinetics for model proteins by purified *S. cerevisiae* proteasome. The proteins consist of four ubiquitin domains fused in series followed by an *E. coli* DHFR domain and different disordered tails at the C terminus. Acidic tail, C-terminal 86 amino acids of Cdc34. 102, 102-amino-acid–long tails derived from *S. cerevisiae* cytochrome \(b_2\). Where no tail is indicated, the protein ended with the C terminus of DHFR. The graph plots the amount of protein estimated by electronic autoradiography in SDS-PAGE gel bands (shown on the right) over time as a percentage of the initial protein amount as described in Online Methods. Data points represent mean values determined from three repeat experiments; error bars indicate s.e.m.
Next we asked whether the acidic region has a dominant effect on the stability of Cdc34 and protects it from degradation even in the presence of other sequences at which the proteasome could initiate degradation. To test this possibility, we attached two tails of different lengths and amino acid sequences to the C terminus of Cdc34 and tracked the proteins’ degradation by yeast proteasome (Fig. 3a). One tail was a 95-amino-acid sequence derived from S. cerevisiae cytochrome b₅; the other tail consisted of 39 amino acids and was derived from an internal region of the E. coli lac repressor. We synthesized Cdc34 with and without the tails by in vitro transcription and translation in E. coli extract, purified the proteins, and induced them to autoubiquitinate by incubating them with ubiquitin, ATP and the E1 enzyme Ube1 (Fig. 3b). Wild-type Cdc34, despite its extensive ubiquitination, remained stable when presented to purified yeast proteasome, but Cdc34 with either of the two tails was degraded (Fig. 3a). We concluded that the proteins were degraded by the proteasome because degradation was inhibited when we added the proteasome inhibitor MG132 or when we depleted ATP in the reaction (Supplementary Fig. 1). Thus, the acidic region is not a dominant protective factor against proteasomal degradation in Cdc34. Instead, it appears that Cdc34 escapes degradation simply because it lacks an effective initiation region.

The acidic region in Cdc34 is characterized by a large number of aspartate and glutamate residues, and it is possible that its high negative charge density prevents proteasome binding. Thus, we tested whether sequences with differently biased amino acid compositions and without the high charge density support degradation. We constructed Cdc34 variants by attaching tails rich in asparagine residues (Cdc34-NRR) or in serine residues (Cdc34-SRR) to Cdc34’s C terminus and investigated proteasomal degradation (Fig. 3a). Both Cdc34 variants underwent ubiquitination (Fig. 3b) but escaped degradation, similar to wild-type Cdc34 (Fig. 3a). Thus, Cdc34 resists degradation because the proteasome is unable to engage it effectively at its disordered region.

Cdc34 is stable in vivo due to poor proteasomal initiation

The experiments described above suggest that the amino acid sequence of the initiation region can affect proteasomal degradation in vitro. In vivo, other factors may affect the process; therefore, we monitored the stability of Cdc34 with and without tails in yeast. We added a hemagglutinin (HA) tag to the N terminus of the Cdc34 proteins described above and expressed them from a centromeric plasmid under the control of a constitutive promoter. We then monitored the accumulation of the different Cdc34 fusion proteins at steady state in mid–log phase by anti-HA western blotting (Fig. 4).

The Cdc34 fusions accumulated at the levels we expected from the in vitro experiments. Western blotting easily detected HA-tagged wild-type Cdc34 (Fig. 4) but barely detected the Cdc34 proteins with tails that served as proteasome initiation sites in vitro (Cdc34-NRR and Cdc34-SRR) unless the proteasome was inhibited by MG132 (Fig. 4). In contrast, Cdc34-NRR accumulated to similar levels as wild-type Cdc34 (Fig. 4). The 95-amino-acid tail did not increase the extent of detectable ubiquitination of Cdc34-95 compared with that of Cdc34 (Supplementary Fig. 2). Thus, in vivo, just as in vitro, the amino acid sequence of the region in Cdc34 at which the proteasome initiates degradation affects the stability of the protein.

Proteasomal initiation of degradation on model substrates

To test whether the amino acid sequence effects are specific to Cdc34 or reflect general preferences by the proteasome for its initiation region, we again monitored degradation of the Ubi₄-DHFR model substrates, but we gave them different C-terminal initiation regions or tails (Fig. 5a). We tested 15 sequences derived from different proteins, most of which are not known to be involved in protein degradation by the UPS (Supplementary Table 2). We synthesized radiolabeled substrates by coupled transcription and translation in E. coli extract and presented them to purified yeast proteasome, which degraded the proteins with very different rates—some proteins remained stable over the entire reaction, whereas others degraded with half-times of minutes (Supplementary Fig. 3). All the tails appeared to be largely disordered, as tested by their sensitivity to a nonspecific protease and judged by CD spectrophotometry (Supplementary Fig. 4). The degradation rates did not correlate with chemical properties of the tails such as total charge, net charge, hydrophobicity, helical propensity, disorder prediction score and side chain volume. Instead, the degradation rates appeared to correlate well with the bias of the amino acid sequence of the initiation region.
We synthesized peptides corresponding to the simplest regions in ten of the tails, presented them to the activated proteasome and monitored proteolysis with the amine-reactive dye fluorescamine, which detects the $\alpha$-amino groups produced during hydrolysis (Supplementary Fig. 5). The proteasome degraded all tails well, and the proteolysis rates did not correlate with the complexity of their amino acid composition (Supplementary Fig. 5). Some of the biased sequences such as those of the SRR and SP2 peptides did not allow the proteasome to initiate degradation (Fig. 5 and Supplementary Fig. 3) but were digested efficiently by the proteasome in the peptide proteolysis assays (Supplementary Fig. 5). Thus, we propose that the proteasome has pronounced preferences for the amino acid sequence at which it initiates degradation.

Amino acid sequence bias affects protein abundance globally

Protein regions with biased amino acid sequences are often disordered; however, disordered regions do not always show biased amino acid composition (Supplementary Fig. 5). Therefore, we asked whether amino acid sequence bias correlates with cellular protein stability globally on a genomic scale.

We previously found that mouse proteins with disordered regions longer than 30 amino acids at either the N or C terminus have shorter lifetimes than proteins without disordered tails by comparing the experimentally determined half-lives of 4,502 mouse proteins with predicted disorder at their N and C termini. The reason for these relationships is presumably that the proteasome requires the disordered regions to engage these proteins and initiate degradation. We have

**Figure 5** Proteasomal preferences for the amino acid composition of initiation regions. (a, b) Plots of in vitro initial degradation rates of model proteins with different C-terminal tails by purified *S. cerevisiae* proteasome as a function of different properties of the tails. The proteins consisted of four ubiquitin moieties fused in series followed by an *E. coli* DHFR domain and a disordered tail at the C terminus. (a) Initial rates of degradation (Supplementary Fig. 3) were plotted against total charge (linear fit, $R = 0.46$), net charge (linear fit, $R = 0.37$), disorder (calculated with DISOPRED2 (ref. 64); linear fit, $R = 0.43$), hydrophobicity (GRAVY scale, ProtParam in ExPASy; linear fit, $R = 0.59$), helix propensity (calculated with Agadir (http://agadir.crg.es))$^{22}$; linear fit, $R = 0.10$) and volume per amino acid (linear fit, $R = 0.27$). Error bars indicate s.e.m. (b) Initial degradation rates for model proteins with each peptide tail were plotted against the amino acid sequence complexity calculated by a SEG algorithm (Supplementary Fig. 3) (sigmoidal fit, $R = 0.90$; linear fit, $R = 0.70$). The solid line is a fit of the sigmoid function to the data. Data points represent means of $n$ measurements, and error bars indicate s.e.m. NB, $n = 4$; NS2, $n = 4$; SNS, $n = 4$; SPmix, $n = 4$; SP1, $n = 3$; SP2, $n = 3$; GRR, $n = 3$; SRR, $n = 3$; DRR, $n = 3$; PEST, $n = 3$; RPB, $n = 3$; eRR, $n = 3$; ODC, $n = 5$; 35\AA K, $n = 6$; Su9, $n = 3$. The amino acid sequence for the various peptides is given in Supplementary Table 2.
Figure 6 Amino acid bias in N- and C-terminal disorder stabilizes proteins in mouse cells. (a,b) Boxplots showing the distribution of protein half-lives in mouse cells grouped by the length and amino acid composition of intrinsically disordered segments at the N terminus (a) and C terminus (b). Proteins were classified into short disordered (<30 amino acids are disordered; gray box), long disordered without amino acid bias (>30 amino acids; light green box) and long disordered with amino acid bias (>30 amino acids; dark green box). The vertical solid black line represents the median, and the boxes represent the first and third quartiles. The notches correspond to ~95% confidence interval for the median. Whiskers connected to the boxes by the dashed lines show the data points up to 1.5× the interquartile range from the box. Points beyond the whiskers were considered outliers and not shown. The number of data points for each group (n), differences between the half-life medians (in hours) of two compared groups (ΔH) and P values indicating the significance in differences in half-life distributions are shown. Half-lives were modeled from relative abundance measurements over time, but estimated stability differences between proteins were robust because the same model was applied to all proteins. Statistical significance of differences was assessed using the Wilcoxon rank-sum test, which is nonparametric and thus does not assume any particular property of the distribution of the data.

Biased amino acid sequences in disease

Htt, the protein associated with HD, has a strikingly biased amino acid composition, and this bias is associated with the etiology of HD. The intensity of the HD phenotype in mouse models increases with the accumulation of a protein fragment that corresponds to exon 1 with 34 or 52 glutamines slowly, if at all; however, attaching a 95-amino-acid–long polyQ region to the N terminus of Htt exon 1 as a protease-binding tag and then presented it to yeast proteasome. The proteasome degraded ubiquitin-tagged Htt exon 1 protein containing 34 or 52 glutamines slowly, if at all; however, attaching a 95-amino-acid–long tail derived from S. cerevisiae cytochrome b2 to the protein C terminus greatly accelerated its degradation (Fig. 7b). Htt exon 1 protein tends to aggregate, but the differences in degradation were not caused by different aggregation states because the proteins were largely, if not completely, monomeric under the assay conditions (Supplementary Fig. 7).

We then tested how well the glutamine and proline regions of Htt exon 1 can support proteasomal initiation individually using the ubiquitin-DHFR model proteasome substrates described above. Neither the 52-amino-acid–long polyQ region (Fig. 7c) nor the PRR...
Biased sequences may bind less tightly to their receptor if they form fewer matching interactions (Fig. 7d) allowed the proteasome to initiate degradation. In contrast, the proteasome rapidly degraded the constructs when we attached a 95-amino-acid–long disordered tail after the polyQ region or PRR (Fig. 7c,d). The proteasome also degraded the constructs when we replaced the polyQ region or PRR with a complex (i.e., unbiased) sequence of similar length (a 50-amino-acid sequence derived from S. cerevisiae cytochrome b6), showing that 50-amino-acid tails are, in principle, able to support degradation. Again, the glutamine-containing substrates remained soluble in these assays and did not aggregate (Supplementary Fig. 7). Thus, it appears that the amino acid sequence of Htt exon 1 is a poor initiation site for the proteasome.

**DISCUSSION**

Most proteins are targeted to the proteasome by a ubiquitin tag, and their degradation also requires an unstructured or disordered region in the substrate, which we call the initiation site or region, and at which the proteasome engages its substrate and initiates degradation. This region has to be located near the ubiquitin tag on the substrate, and it has to be long enough to allow the substrate to access the proteasome’s degradation machinery. Here we propose that the amino acid sequence of the initiation region also affects degradation and that the proteasome has distinct preferences for the site at which it engages and binds its substrate to initiate degradation. In particular, it seems that the proteasome can engage proteins inefficiently at polypeptide sequences with biased amino acid compositions and that very biased sequences can escape proteasomal initiation entirely.

Researchers first discovered that amino acid repeat sequences are associated with unexpected stability in the Epstein-Barr virus protein EBNA1 (ref. 22). Epstein-Barr virus infects B lymphocytes and forms stable episomes whose maintenance requires the virally encoded protein EBNA1 (ref. 42), which is protected from proteasomal degradation by glycine-alanine repeats. Biased amino acid sequences also reduce the processivity of the proteasome and cause the production of partially degraded protein fragments by stalling the progression of the proteasome along its substrate, but the molecular mechanism of these effects is not known. How can biased amino acid sequences impede substrate engagement by the proteasome? One possibility is that amino acid composition is directly related to the binding affinity of the initiation region to the proteasome. Because the proteasome must bind and degrade many different substrates, it is likely that the proteasome’s receptor for the initiation regions recognizes several patterns of chemical features (hydrophobicity, charge, etc.) that are complementary to the receptor surface rather than one strict consensus sequence (Fig. 8a). A diverse set of sequences would then be able to bind the receptor well enough to serve as initiation sites for the proteasome. In other words, the threshold for serving as a good initiation sequence would be achieved by partial matches between the receptor surface and the initiation site sequence (Fig. 8b). Under these circumstances, the likelihood of an amino acid sequence binding to the proteasome above any given threshold would correlate with the complexity of the sequence: biased sequences would be less likely to satisfy the required interactions with the receptor surface than complex sequences. Thus, the correlation between initiation of degradation and sequence complexity could reflect specific sequence-binding preferences by the proteasome.

A second possibility is that amino acid composition affects the structure and compactness of the initiation site and thus its ability to reach its receptor on the proteasome. Disordered polypeptide sequences adopt a diverse set of conformations that are in equilibrium with each other. Amino acid composition affects the conformational ensemble and may therefore affect the ability of an initiation region to reach its receptor on the proteasome (Fig. 8c).

Our biochemical experiments tested only 15 different sequences, so the relationship between amino acid sequence composition and degradation cannot be extrapolated to make general predictions. Nevertheless, the relationship is reflected in the three different natural proteins we investigated. In all three proteins, the disordered regions that are available as proteasome initiation sites have biased amino acid sequences. Interestingly, inspection of the linker regions in the UbL-UBA proteins suggests that the property of bias is conserved between UbL-UBA proteins but that the nature of the bias is not: different amino acids are over-represented in different linkers and in different UbL-UBA proteins. Even more strikingly, the stability of soluble proteins in mouse cells correlates with the amino acid sequence complexity in the disordered regions in these proteins. Thus, the proteasome’s amino acid sequence preferences seem to globally fine-tune protein turnover. Disordered regions in proteins typically evolve more rapidly and with fewer restraints than the regions that form folded domains, presumably because changes in disordered regions are less likely to affect the structure of a folded region and hence the molecular function mediated by the folded domain. Therefore, evolving the sequence composition of a disordered region of a protein provides a simple genetic mechanism to change protein levels without directly affecting the molecular function of a protein.

For any one specific protein, its stability will depend on the relationship between the ubiquitin modifications and initiation region. For example, in the UbL-UBA protein Rad23, the three linkers connecting UbL and UBA represent the only possible initiation regions for the

---

**Figure 8** Amino acid sequence composition bias may directly affect recognition of initiation sites by the proteasome. (a–c) Schematic representation of initiation sites of different amino acid compositions being recognized by the 26S proteasome to illustrate two different models for the relationship between sequence complexity and proteasome recognition. The 26S proteasome is represented by gray shapes; ubiquitin by green spheres; the folded domain of the substrate by the large yellow sphere; and the initiation region by the black, red, blue, yellow and green tails (a). Biased sequences may bind less tightly to their receptor if they form fewer matching interactions than complex sequences (b). Alternatively, biased sequences may form compact rigid structures and access their binding site in the degradation channel less easily than flexible disordered regions (c).
proteasome. The amino acid compositions of the linkers are clearly biased, so it is straightforward to relate sequence composition to degradation by the proteasome. In other proteins, only one particular stretch of their disordered regions may have a biased amino acid composition, and therefore the proteasome can initiate degradation elsewhere within the protein. Thus, to be able to predict the effect of biased sequences on protein degradation more broadly, we need to understand where proteins are ubiquitinated and where exactly the proteasome initiates degradation relative to the ubiquitination site. It is also possible that mechanisms exist to create disordered regions throughout proteins, e.g., when ubiquitination of a folded domain leads to its unfolding52. Finally, it is quite possible that accessory factors such as Cdc48 (p97) make the proteasome less dependent on the presence of initiation regions for degradation53,54.

The proteasome's struggle to initiate degradation at biased amino acid sequences may contribute to the accumulation of some proteins that form aggregates associated with neurodegenerative diseases. We have found that, at least in vitro, the proteasome struggles to initiate degradation of the Htt fragment that accumulates in HD. Proteins related to other neurodegenerative diseases also carry abnormal glutamine repeats in their disease-associated forms55. Examples are atrophin-1, which is associated with dentatorubralpallidoluysian atrophy; androgen receptor, which is associated with spinobulbar muscular atrophy; and ataxin proteins, which are associated with spinocerebellar ataxia. The neuronal aggregates of these proteins contain ubiquitin, yet the proteasome apparently fails to degrade them56. Similarly, tau protein forms aggregates called neurofibrillary tangles in neurons of people with Alzheimer’s disease, and the protein in the tangles is ubiquitinated56,57. Tau protein contains long stretches largely made up of five amino acids; thus, it is possible that tau accumulates because the proteasome struggles to initiate degradation at the repeat sequences.

The Htt exon 1 fragment is made up almost entirely of glutamine and proline repeats, and therefore no other sequences are available for proteasomal initiation. However, this is not the case for the other proteins listed. To know if and how the initiation site sequence contributes to its stability, it will be necessary to map the ubiquitination sites and determine how the proteasome selects its initiation sites. Other mechanisms likely contribute to the failure of proteasome degradation55,58–61, and different mechanisms of Htt accumulation can demand opposing therapeutic strategies. If Htt is not recognized by the proteasome, interventions that enhance the interaction between the two are needed. Indeed, enhancing proteasome activity can reduce the accumulation of disease-associated protein aggregates62,63.

On the other hand, if Htt clogs the proteasome, it may be beneficial to weaken the interaction. Thus, it will be important to determine which mechanism is physiologically relevant.

In summary, we found that the proteasome has pronounced preferences for the amino acid sequence of the substrate at the site at which it initiates degradation. These preferences affect substrate selection and may represent a second component of the proteasome-targeting code superimposed on the ubiquitination code. Amino acid sequence variation within disordered regions can affect cellular protein half-life without directly affecting molecular function and may be a general genetic mechanism that has implications in linking genotype to phenotype.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Brickner and R.A. Lamb (Northwestern University), M. Glickman (Technion), R.V. Pappu (Washington University) and Y. Saeki (Tokyo Metropolitan Institute of Medical Science) for advice and reagents, as well as A. Zolkarkar for early experiments on this project. We are also grateful to K. Brown, A. Gnanam and E.J. Cho for help with the CD experiments. This work was supported by the US National Institutes of Health (ROI GM063004, U54GM105816 and U54CA143869 (A.M.)); the Welch Foundation (F-1817 (A.M.)); the Program to Disseminate Tenure Track System from the Ministry of Education, Culture, Sports, Science and Technology, Japan (T.I.); the UK Medical Research Council (MC, U105185859 (S.C. and M.M.B.); and the European Molecular Biology Organization (Long-Term Fellowship (S.C.) and Young Investigator Program (M.M.B.).)

AUTHOR CONTRIBUTIONS

S.F., T.I., E.L., H.Y. and G.K. performed experiments, analyzed data and co-wrote the paper. M.M.B and A.M. directed experiments, analyzed data and co-wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Schrader, E.K., Harstad, K.G. & Matuschek, A. Targeting proteins for degradation. Nat. Chem. Biol. 5, 815–822 (2009).
2. Komander, D. & Rape, M. The ubiquitin code. Annu. Rev. Biochem. 81, 203–229 (2012).
3. Prakash, S., Tian, L., Ratliff, K.S., Leholtzky, R.E. & Matuschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat. Struct. Mol. Biol. 11, 830–837 (2004).
4. Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M. & Matuschek, A. ATP-dependent proteases degrade their substrates by processively unrolling them from the degradation signal. Mol. Cell 7, 627–637 (2001).
5. Nathan, J.A., Kim, H.T., Ting, L., Gogi, S.P. & Goldberg, A.L. Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? EMBO J. 32, 552–565 (2013).
6. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S. & Matuschek, A. Rad23 escapes degradation because it lacks a proteasome initiation region. Nat. Commun. 2, 192 (2011).
7. Inobe, T., Fishbain, S., Prakash, S. & Matuschek, A. Defining the geometry of the two-component proteasome degron. Nat. Chem. Biol. 7, 161–167 (2011).
8. Schauer, C. et al. Rad23 links DNA repair to the ubiquitin/proteasome pathway. Nature 391, 715–718 (1998).
9. Watkins, J.F., Sung, P., Prakash, L. & Prakash, S. The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. Mol. Cell. Biol. 13, 7757–7765 (1993).
10. Heesen, S., Masucci, M.G. & Dantuma, N.P. The UBZ2 domain functions as an intrinsic stabilization signal that protects Rad23 from proteasomal degradation. Mol. Cell 18, 225–235 (2005).
11. Heesen, C., Acs, K., Hoogstraten, D. & Dantuma, N.P. C-terminal UBZ domains prevent ubiquitin receptors by preventing initiation of protein degradation. Nat. Commun. 2, 191 (2011).
12. Punta, M. et al. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301 (2012).
13. Stack, J.H., Whitney, M., Rodens, S.M. & Pollok, B.A. A ubiquitin-based tagging system for controlled modulation of protein stability. Nat. Biotechnol. 18, 1298–1302 (2000).
14. Goebl, M.G., Goetsch, L. & Byers, S. The Ubuc3 (Cdc34) ubiquitin-conjugating enzyme is ubiquitinated and phosphorylated in vivo. Mol. Cell. Biol. 14, 3022–3029 (1994).
15. Banerjee, A., Gregori, L., Xu, Y. & Chau, V. The bacterially expressed yeast CDC34 gene product can undergo autoubiquitination to form a multibuquitin chain-linked enzyme. J. Biol. Chem. 268, 5668–5675 (1993).
16. Chau, V. et al. A multibuquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243, 1576–1583 (1989).
17. Bachmair, A. & Varshavsky, A. The degradation signal in a short-lived protein. Cell 56, 1019–1032 (1989).
18. Wotton, J.C. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput. Chem. 18, 269–285 (1994).
19. Wotton, J.C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. Methods Enzymol. 266, 554–571 (1996).
20. Daskalogianni, C. et al. Gly-Ala repeats induce position- and substrate-specific regulation of 26 S proteasome-dependent partial processing. J. Biol. Chem. 283, 30090–30100 (2008).
21. Hoyt, M.A. et al. Glycine-alanine repeats impair proper substrate unfolding by the proteasome. EMBO J. 25, 1720–1729 (2006).
22. Sharipio, A., Imreh, M., Leochiks, A., Imreh, S. & Masucci, M.G. A minimal glycin-ala-tine repeat prevents the interaction of ubiquitinated IxB with the proteasome. Nat. Med. 4, 939–944 (1998).

23. Zhang, M. & Coffino, P. Repeat sequence of Epstein-Barr virus-encoded nuclear antigen 1 protein interrupts proteasome substrate processing. J. Biol. Chem. 279, 8635–8641 (2004).

24. Tian, L., Holmgren, R.A. & Matouschek, A. A conserved processing mechanism regulates the activity of transcription factor Ctsib interruptus and NF-kB. Nat. Struct. Mol. Biol. 12, 1045–1053 (2005).

25. Pratt, G. & Rechsteiner, M. A 4A consensus proteasome inhibitor on. Nature 407, 818–822 (2000).

26. Tian, L., Holmgren, R.A. & Matouschek, A. A conserved proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 278, 871–872 (1997).

27. Yang, D. & Jin, Y. A proteasome inhibitor on. J. Biol. Chem. 272, 1791–1798 (1997).

28. Udrenfri, S. et al. Fluorescine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science 182, 871–872 (1972).

29. Saha, K. & Saha, S. A proteasome inhibitor on. Proteins 4, 38–48 (2001).

30. Tompa, P. & Fersht, A.R. Structure and Function of Intrinsically Disordered Proteins (CRC Press, 2010).

31. Jain, D. & Lee, H. et al. Intrinsically disordered segments affect protein half-life in the cell and during evolution. Cell Rep. 8, 1832–1844 (2014).

32. Schwahn, C., B. et al. Global quantification of mammalian gene expression control. Nature 473, 337–342 (2011).

33. Zucca, C., Valenza, M. & Cattaneo, E. Molecular mechanisms and potential therapeutic targets in Huntington's disease. Physiol. Rev. 90, 905–981 (2010).

34. Diffiglia, M., Sapp, E., Chase, K.O., Davies, S.W. & Bates, G.P. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 277, 1990–1993 (1997).

35. Kalchman, M.A. et al. Huntington is ubiquitinated and interacts with a specific ubiquitin-conjugating enzyme. J. Biol. Chem. 271, 19385–19394 (1996).

36. Douglas, P.M. & Dillin, A. A proteasome inhibitor on. J. Cell Biol. 190, 719–729 (2010).

37. Jain, N.R., Zemskov, E.A., Gh, W. & Nukina, N. Impaired proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin induces apoptosis by caspase activation through mitochondrial cytochrome c release. Hum. Mol. Genet. 10, 1049–1059 (2001).

38. Wyttenbach, A. et al. Accumulation of mutant huntingtin fragments in aggresome-like inclusions: a role for their classification. Mol. Cell. Biol. 28, 7068–7078 (2013).

39. Venkataraman, R., Wetzel, R., Tanaka, M., Nukina, N. & Goldberg, A.L. Urykatory proteases cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. Mol. Cell 14, 95–104 (2004).

40. Zoghbi, H.Y. & Orr, H.T. Glutamine repeats and neurodegeneration. Annu. Rev. Neurosci. 23, 217–247 (2000).

41. Wyttenbach, A. et al. An antigenic profile of Lewy bodies: immunocytochemical indication for protein phosphorylation and ubiquitination. J. Neuropathol. Exp. Neurol. 48, 81–93 (1989).

42. Iqbal, K. & Grundlehner, I. Ubiquitination and phosphorylation of paired helical filaments in Alzheimer's disease. J. Neurobiol. 5, 399–410 (1991).

43. Kaelin, N. & Frank, A.H. et al. Development of the multiple sequence approximation within the human proteasome. EMBO J. 23, 4307–4318 (2004).

44. Kraut, D.A. & Matouschek, A. A proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 278, 871–872 (1997).

45. Kraut, D.A. et al. Structure and species-dependence of proteasomal processivity. ACS Chem. Biol. 7, 1444–1453 (2012).

46. Babu, M.M., Kriwacki, R.W. & Pappu, R.V. Structural biology. Versatility from protein design. Science 337, 1460–1461 (2012).

47. Mao, J.-H., Cricci, S.L., Vitalis, A., Chicoine, C.L. & Pappu, R.V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 107, 8183–8188 (2010).

48. Das, K.R. & Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc. Natl. Acad. Sci. USA 110, 13392–13397 (2013).

49. Müller, S. et al. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 107, 14609–14614 (2010).

50. Moes, H.A., Wakabayashi, S., Nakai, K. & Patil, A. Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. Mol. Biosyst. 8, 3262–3273 (2012).

51. Brown, C.J., Johnson, A.K., Dunker, A.K. & Daughdrill, A.W. Evolution and disorder. Curr. Opin. Struct. Biol. 21, 441–446 (2011).

52. Hagai, T. & Levy, Y. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. Proc. Natl. Acad. Sci. USA 107, 2001–2006 (2010).

53. Beskow, A. et al. A conserved unfolded activity for the p97 AAA-ATPase in proteasomal degradation. J. Mol. Biol. 394, 732–746 (2009).

54. Raman, M., Haven, C.G., Walter, J.C. & Harper, J.W. A genome-wide screen identifies p97 as an essential regulator of DNA damage-dependent CDT1 destruction. Mol. Cell 44, 72–82 (2011).

55. Zoghbi, H.Y. & Orr, H.T. Glutamine repeats and neurodegeneration. Annu. Rev. Neurosci. 23, 217–247 (2000).

56. Bancher, C. et al. An antigenic profile of Lewy bodies: immunocytochemical indication for protein phosphorylation and ubiquitination. J. Neuropathol. Exp. Neurol. 48, 81–93 (1989).

57. Iqbal, K. & Grundlehner, I. Ubiquitination and phosphorylation of paired helical filaments in Alzheimer's disease. J. Neurobiol. 5, 399–410 (1991).

58. Bence, N.F., Sampat, R.M. & Kopito, R.R. Impairment of the ubiquitin-proteasome system by protein aggregation. Science 292, 1552–1555 (2001).

59. Hopp, M.S. et al. Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease. J. Cell Biol. 196, 573–587 (2012).

60. Holmgren, R.A. & Matouschek, A. E. & Morimoto, R.I. Inefficient degradation of truncated polyglutamine proteins by the proteasome. EMBO J. 23, 4307–4318 (2004).

61. Bennett, E.J., Bence, N.F., Jayakumar, R. & Kopito, R.R. Global impairment of the ubiquitin-proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion body formation. Mol. Cell 17, 351–365 (2000).

62. Kruegel, U. et al. Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLOS Genet. 7, e1002253 (2011).

63. Lee, B.H. et al. Enhancement of proteasomal function by a small-molecule inhibitor of USP14. Nature 467, 179–184 (2010).

64. Ward, J.J., Siddhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. 337, 635–645 (2004).

65. Mulloz, V. & Serrano, L. Development of the multiple sequence alignment within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41, 495–509 (1997).
ONLINE METHODS

Substrate proteins. Substrates were derived from S. cerevisiae cytochrome b₂, Cdc34 or Rad23, E. coli DHFR and Homo sapiens huntingtin. Their coding sequences were cloned either into the plasmid pGEM-3Z (+) (Promega) for in vitro expression or into the yeast CEN plasmid p416 GDP for in vivo experiments as described previously66.

N-terminal ubiquitin tags were composed of four copies of the coding region for ubiquitin, each containing the mutation Gly76 to Val and connected to the next ubiquitin by the linker sequence Gly-Ser-Gly-Gly-Gly as described previously13. C-terminal tails derived from cytochrome b₂ sequences were attached to DHFR and the other proteins through a short linker, and lysine residues in the tails were replaced by arginine. The tail sequences are given in Supplementary Tables 1 and 2. Ub₄-DHFR-15, Ub₂-DHFR-64, Ub₂-DHFR-102, Ub₄-DHFR-34Q-102 and Ub₂-DHFR-52Q-102 all contained a hexaarginine tag on the C-terminus; Ub₂-DHFR-50 did not. In Ub₂-DHFR-PRR, the tail consisted of the prion-like region of Htt exon 1 (residues 51–100 in the Htt exon 1 variant with a 34-residue glutamine repeat), followed by residues 1–95 from cytochrome b₂ where indicated. The tail of Ub₂-DHFR-acidic tail contained the last 86 amino acids of Cdc34. In Ub₂-DHFR-L1, Ub₂-DHFR-L2 and Ub₂-DHFR-L3, the tails consisted of the Rad23 linker regions as follows: L1 connects the Ubl and the N-terminal UBA domains and corresponds to amino acids 77–144 of Rad23; L2 connects the N-terminal UBA and the Rad4-binding domain and corresponds to amino acids 186–250 of Rad23; and L3 connects the Rad4-binding domain and the C-terminal UBA domains and contains amino acids 286–355 of Rad23.

The constructs containing Htt exon 1 consisted of the N-terminal four-ubiquitin tag described in the preceding paragraph followed by the sequence of Htt exon 1 with either 34 or 52 residue glutamine repeats, i.e., amino acids 1–90 (34Q) or 1–109 (52Q) of Htt. The Htt sequence was followed by amino acids 1–50 or 1–95 of cytochrome b₂ as indicated.

The Cdc34 constructs in Figure 2 consisted of the entire CDC34 coding sequence followed by the tails described below, and the constructs expressed in yeast included an N-terminal HA tag. The C-terminal tails consisted of residues 321–354 of the E. coli lac repressor but with lysine residues replaced by arginines (Cdc34-39), four copies of residues 373–386 from transcription factor Spț2 (Cdc34-NRR), four copies of residues 178–196 of transcription factor ICP4 (Cdc34-SRR), or residues 1–95 of S. cerevisiae cytochrome b₂. Tail sequences are given in Supplementary Tables 1 and 2.

In vitro autoubiquitination. In vitro translated radioiodelabeled substrates were ubiquitinated at 30 °C for 16 h in a reaction mixture containing 25 mM Tris-HCl (pH 7.4), 0.01% SDS, 1 mM DTT, 1 mM ATP, 10 mM creatine phosphate, 0.1 mg/ml creatine phosphokinase. 250–μM (final concentration) peptides were added to purified proteasome in reaction buffer to start the proteolysis. Samples were withdrawn at the indicated times, added to equal volume of 10% TCA and incubated for 5 min at 65 °C. The mixtures were neutralized by 25 volumes of 0.2 N NaHPO₄ on ice. Finally, 1/40 volume of 25 mg/mL fluorescamine in DMSO was added and mixed vigorously. Fluorescence was measured using an excitation wavelength of 390 nm and an emission of 475 nm. Peptides were custom synthesized (Genscript) apart from peptides NB, NS2, SP1, SP2 and SP231, which were gifts from R.A. Lamb (Northwestern University).

Protein expression and purification. Yeast proteasome was purified from S. cerevisiae strain YYS40 (MATa rpn11::RNPl1 13 ‧ Flag-HIS3 leu2 his3 p1 ade2 can1 sld1) by immunofinity chromatography using FLAG antibodies (M2 agarose affinity beads, Sigma) as described previously with modifications66. Protein preparations were analyzed by SDS-PAGE and compared to published compositions67. A typical gel with assigned bands is shown in Supplementary Figure 8. Each proteasome preparation was checked for activity by testing degradation of the proteasome substrate Ub₄-DHFR-95 and for contamination by proteases by testing for stability of proteins that lack a proteasome-binding tag (DHFR-95).

For in vitro degradation experiments, radioactive substrates were expressed from a T7 promoter by a coupled in vitro transcription–translation reaction using E. coli T7 S30 Extract System for Circular DNA (Promega) containing [35S]-methionine following the manufacturers protocol. Htt substrates were expressed from a T7 promoter by in vitro transcription–translation using the RTS 100 E. coli HY Kit (Roche) containing [35S]-methionine according to the manufacturer's instruction in 25-μL reactions. After synthesis, the substrates were either partially purified by high-speed centrifugation followed by precipitation in two volumes of saturated (NH₄)₂SO₄ or affinity purified using Talon magnetic beads (Clontech) as described previously66.

Proteasomal degradation assays. Degradation assays were performed as described previously66. Briefly, assays were carried out at 30 °C by adding radioiodelabeled substrates to 50 nM of purified yeast proteasome in a reaction buffer containing a creatine-phosphate creatine kinase ATP-regenerating system. Samples were removed at designated times, added to SDS-PAGE sample buffer to stop the reaction and analyzed by SDS-PAGE. Protein amounts were determined by electronic autoradiography (Instant Imager, Packard). Each assay was repeated at least three times. Initial degradation rates are given by the slope of the decay curves at time zero in Supplementary Figure 3 and are calculated as the product of the amplitude and the rate constant of the decay curve determined by nonlinear fitting to a single exponential decay in the software package Kaleidagraph (version 4.1, Synergy Software). Original images of autoradiograms can be found in Supplementary Data Set 1.

In vivo protein abundance determination. Cdc34 fusion proteins were expressed under the control of the constitutive glyceraldehyde 3-phosphate dehydrogenase (GPD) promoter from a CEN plasmid (p416 GPD) with a URA3 selection marker in S. cerevisiae strain BY4741 pdr5A (MATa his3Δ1 leu2α met15α ura3Δ3 pdr5::kanMX4). Cells were grown to mid–log phase and lysed by vortexing with glass beads (BioSpec Products). Protein extracts were prepared and analyzed by western blotting using standard protocols as described66. Cdc34 protein was detected with a monoclonal anti-HA antibody (1:5,000, Sigma, #H9658) and an Alexa-800–labeled goat anti-mouse secondary antibody (1:20,000, Rockland Immunochemicals, #610-132-121), and Ssc2p was detected by an anti-SCS2 polyclonal antibody (1:1,000, gift from J. Brickner (Northwestern University)) and an Alexa-680 goat anti-rabbit secondary antibody (1:20,000, Invitrogen, #A21109). Protein amounts were estimated by direct infrared fluorescence imaging (Odysee LICOR Biosciences). Original images of western blots can be found in Supplementary Data Set 1.

Peptide proteolysis assay. Peptide proteolysis assays were carried out according to the method of Evans and Ridella66. Proteolysis reactions were performed with 200 nM purified yeast proteasome at 30 °C in 5% (v/v) glycerol, 5 mM MgCl₂, 50 mM Tris-HCl (pH 7.4), 0.01% SDS, 1 mM DTT, 1 mM ATP 10 mM creatine phosphate, 0.1 mg/ml creatine phosphokinase. 250-μM (final concentration) peptides were added to purified proteasome in reaction buffer to start the proteolysis. Samples were withdrawn at the indicated times, added to equal volume of 10% TCA and incubated for 5 min at 65 °C. The mixtures were neutralized by 25 volumes of 0.2 N NaHPO₄ on ice. Finally, 1/40 volume of 25 mg/mL fluorescamine in DMSO was added and mixed vigorously. Fluorescence was measured using an excitation wavelength of 390 nm and an emission of 475 nm. Peptides were custom synthesized (Genscript) apart from peptides NB, NS2, SP1, SP2 and SP231, which were gifts from R.A. Lamb (Northwestern University).

Determination of the global effect of amino acid bias on protein stability. Mouse protein half-life data was obtained from Schwanhäusser et al.61. Intrinsically disordered was predicted for all studied protein sequences (downloaded from UniProtKB/Swiss-Prot, http://www.uniprot.org/) using three complementary methods: DISOPRED2 (ref. 64), IUPRED long66 and PONDVR V15 (ref. 70). Based on the length of the disordered segments, proteins were first classified as those that had short (stretches of ≤30 disordered residues) and long (stretches of >30 disordered residues) for N and C termini separately disordered termini. Amino acid bias within N- and C-terminal long disordered segments was identified using LPS-annotate71,72 employing default parameters. Using a stringent detection P value cutoff of <1 × 10⁻¹⁰, the proteins with long disordered termini were classified into those with and without amino acid bias. Statistical significance of the difference in the distributions of half-life values among different classes of proteins was estimated using the nonparametric Wilcoxon rank-sum test.