# Design principles that protect the proteasome from self-destruction

**Amit Kumar Singh Gautam**
The University of Texas at Austin
**Houqing Yu**
The University of Texas at Austin
**Christopher Yellman**
The University of Texas at Austin
**Adrian Elcock**
University of Iowa
**Andreas Matouschek** ( matouschek@austin.utexas.edu )
The University of Texas at Austin https://orcid.org/0000-0001-6016-2341

**Article**

**Keywords:** ubiquitin tag, localization, proteasomal subunits

**Posted Date:** July 8th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-38559/v1

# Abstract

The proteasome establishes protein homeostasis by selectively degrading cellular proteins. There are two main prerequisites for a protein to be selected for proteasomal degradation: a ubiquitin tag covalently attached to the protein and a disordered region within the protein. The ubiquitin tag localizes the protein to the proteasome, which then engages it at the disordered region to feed it into the proteolytic particle for degradation. Here we ask how the proteasome avoids the degradation of its own subunits, which are by definition localized to the proteasome, and thus ensures the integrity of this essential proteolytic machine. We found that though many proteasomal subunits have disordered tails, these tails are composed of amino acid sequences that escape recognition by the proteasome, even when they are not able to reach the translocation channel.


# Introduction

The proteasome is a multi-subunit protease that controls the concentrations of thousands of proteins. It is abundant in the nucleus and cytosol of all eukaryotic cells, where it clears misfolded and damaged proteins as part of the stress response, degrades regulatory proteins to control cellular processes, and processes foreign proteins as part of the adaptive immune response1–3.

The proteasome itself consists of at least 33 different polypeptides that form a 2.5 MDa complex4. It is composed of a cylindrical 20S core particle, which contains the proteolytic sites, and a 19S regulatory particle, which contains ubiquitin-binding sites, enzymatic sites that remove ubiquitin chains, and a ring of ATPase subunits that function as the substrate translocation motor4. The regulatory particles dock on either end of the core particle to form the functional 26S proteasome. The proteolytic sites exist as three pairs, one with chymotrypsin-like sequence preferences, one with trypsin- like preferences, and one with caspase-like preferences5. Working together, these proteases can cleave effectively any amino acid sequence. However, the proteolytic sites are located on an internal surface of the core particle and thus only accessible through a roughly 55 Å long translocation channel that starts at the surface of the regulatory particle and runs along the long axis of the proteasome. Access to the channel is controlled by the ring of ATPase subunits in the regulatory particle and a ring of loops some 35 Å down the channel from its entrance are thought to engage the substrate and drive it to the proteolytic sites.

Proteins are typically targeted to the proteasome by the covalent attachment of several copies of the small protein ubiquitin6,7. The ubiquitin tags in turn are recognized by proteasome subunits that serve as substrate receptors. In many cases the localization of a protein to the proteasome suffices to induce its proteolysis8,9. Degradation initiates when the proteasome engages the substrate at a disordered region in the substrate, translocates its polypeptide chain down the channel to the proteolytic sites in the interior of the complex, unfolding the substrate in the process. Thus, proteins are degraded when they are bound to the proteasome in a manner that allows the proteasome to engage the proteins at a disordered region2,4,10,11.

The proteasome destroys target proteins while it itself remains stable and persists in the cell with a long half-life (e.g.,16 h in immortalized MEFs and about two weeks in rat liver)12–14. Proteasome subunits are localized on the proteasome by definition and many of them contain disordered regions. These regions perform regulatory roles and contain a multitude of post-translational modifications or form binding sites for proteasome interacting proteins (PIPs)15–17. Thus, with the two requirements for proteasomal degradation fulfilled, the question arises how the proteasome subunits themselves escape destruction.

Here we find that the integrity of subunits is ensured by preventing the proteasome from initiating degradation on these polypeptide chains. Some subunits lack disordered regions or their disordered regions are buried within the proteasome complex (or possibly by binding partners). However, where subunits do have exposed disordered regions, they are constructed from an amino acid sequence that fails to be recognized by the proteasome and thus prevents it from initiating degradation. This appears to apply even for disordered regions that are placed too far from the entrance to the degradation channel to be engaged effectively.

# Results

*Several proteasome subunits contain disordered regions*

Proteasomal degradation begins at disordered regions within proteins. Regions within several subunits of the regulatory particle are not defined in the recent high-resolution structures of yeast proteasome (e.g., PDB 6FVT, and 6FVU), suggesting that these regions are disordered18. For example, six of the non-ATPase subunits end in stretches of 20 residues or more that lack a defined structure (Table 1). We first asked which of these tails, if any, were long enough to reach the entrance of the translocation channel and thus could make the proteasome susceptible to accidental self-digestion and destruction.

We built a complete atomic model of the yeast proteasome based on recent high- resolution structures (PDB 6FVT and 6FVU) and added the unresolved residues in the tails using homology modeling. The tail structures were built by assembling amino acids in random order from di-peptide conformations found in a library of non-redundant high- resolution protein structures19,20 (see Methods, Fig. 1A, and Supplementary Data 1). The length of the assembled tails was calculated as the distance between the Cα atom of the last resolved residue at one end of the sequence and the Cα atom at the other end of the tail (Supplementary Fig. 1A). We also calculated the direct through-space distance as the shortest possible connection between the Cα atom of the last resolved residue and entrance to the translocation channel (Supplementary Fig. 1B). The ratio of these two distances provides some measure of the likelihood of a tail reaching the entrance to the translocation channel: the larger the ratio, the more likely the tail can reach the channel entrance, with 1.0 the theoretical threshold, though in reality, steric constraints imposed by the rest of the proteasome structure could make the true threshold somewhat higher than this (Fig. 1B). The C-terminal disordered tails of Rpn1, Rpn2, and Rpn13 are either too short or too distantly located to reach the channel entrance (ratio of max tail length to distance from the

translocation channel < 0.5 for Rpn2, Rpn13, and 0.7 for Rpn1, Fig 1B). On the other hand, the disordered tails of some subunits were either long enough (Rpn10 ratio ≥ 1) or almost so (Rpn3, and Rpn8, ratio ≥ 0.8) to reach the entrance of the translocation channel (Fig. 1B and Table 1).

We chose three subunits with disordered C-terminal tails to investigate further. We chose Rpn10 and 13 because they are substrate receptors and their binding partners become degraded routinely, suggesting that they themselves might be susceptible to degradation. We chose Rpn3 because it is an essential structural component of the proteasome complex and its disordered tail is potentially within reach of the entrance channel. Finally, we focused on C-terminal tails because our experimental tools are better suited for their analysis (see below).

*Rpn13 escapes degradation because it does not access the translocation channel*

The Rpn13 subunit is one of the three ubiquitin receptors; substrates bound to Rpn13 are degraded while the receptor itself escapes. Although the last 23 amino acids of Rpn13 appear to be disordered, the subunit is located apically, near the tip of the 19S regulatory particle, and unlikely to reach the entrance to the substrate channel and the protein translocation motor (Figs. 1, 2A and Table 1). If Rpn13 escaped degradation simply because of its location and failure to engage the motor, extending its C-terminus by an artificial tail should lead to its destruction (Fig. 2B). To determine the abundance of Rpn13 and derivatives, we attached a hemagglutinin (HA) tag to its N-terminus and expressed the protein in yeast under the control of a GAL1 promoter from a CEN/ARS plasmid (Fig. 2B). As expected, HA-Rpn13 was easily detected by western blotting (Fig. 2C). Indeed, attaching an artificial tail with a sequence that can be recognized by the proteasome (Su9,21,22) to the C-terminus of HA-Rpn13 led to its degradation (HA-Rpn13-Su9; Fig. 2C), whereas attaching a tail with a sequence that escapes proteasome recognition (HA-Rpn13-SP2,21,22 ) allowed HA-Rpn13 to accumulate just as HA-Rpn13 without the additional tail (HA-Rpn13; Fig. 2C).

Rpn13 with a suitable extended tail was degraded by the proteasome as it was stabilized by treatment with the proteasome inhibitor MG132. Ubiquitination was not required for degradation, presumably because as a proteasome subunit, Rpn13 is by default localized to the proteasome, even without a ubiquitin tag (Fig. 2C, and Supplementary Fig 2A). Attenuating the first enzyme in the ubiquitination pathway (ubiquitin-activating enzyme Uba1), by using a temperature-sensitive strain23, did not affect the abundance of HA-Rpn13-Su9 (Supplementary Fig. 2A), whereas a substrate with the classic N-end degron24,25, which depends on ubiquitination, was stabilized (Supplementary Fig. 2B).

The HA-tagged Rpn13 mutants were incorporated into the proteasome, at least in part, as they were detectable by western blotting after immunoprecipitation of proteasome particle through a FLAG epitope fused to Rpn11 (Supplementary Fig. 3). Degradation of Rpn13 with a tail (Rpn13-Su9) was not due to its over-expression because integration of Rpn13-Su9 at its genomic locus (FLAG-Rpn13-Su9) and expression from its native promoter also led to its degradation, whereas FLAG-Rpn13 without a tail as well as with a tail that is not recognized by the proteasome accumulated (Fig. 2D).

Thus, Rpn13 escapes degradation despite its location on the proteasome in the vicinity of the entrance to the translocation channel because its unstructured region is too short to reach the entrance to the channel and thus cannot be engaged by the proteasome.

*Rpn10 and Rpn3 are protected by their amino acid sequence*

The Rpn10 is another ubiquitin receptor and again substrates bound to Rpn10 are degraded while Rpn10 itself escapes. Rpn10 also contains a disordered region at its C- terminus and this tail is long enough to reach the entrance of the translocation channel (Fig. 3A, Table 1). This raises the question how Rpn10 escapes degradation.

The proteasome has distinct preferences for the amino acid sequence of the polypeptide at which it initiates degradation. If Rpn10 escaped degradation because its amino acid sequence masks it from recognition, then replacing the relevant region with a sequence that can be engaged by the proteasome should lead to Rpn10's destruction. To test this model, we attached an HA-tag to the N-terminus of Rpn10 and constructed two further hybrid proteins by replacing Rpn10's C-terminal tail (amino acids 243-268) with a sequence recognized by the proteasome (Su921,22) or with a sequence that is ignored (SP221,22) (Fig. 3B). Rpn10 and its derivatives were again expressed from a GAL1 promoter on a CEN/ARS plasmid and their abundance determined by western blotting. HA-Rpn10 with its native tail accumulated but replacing the tail with the Su9 sequence led to its degradation (Fig. 3C). It is unlikely that the transplantation of the tail in itself caused Rpn10's destabilization because the Rpn10 hybrid with a SP2 tail was stable (Fig. 3C). The Rpn10 degradation was by the proteasome and did not require ubiquitination (Fig 3C, and Supplementary Fig. 2A). We observed similar protein levels of HA-Rpn10 mutants when expressed from Rpn10's native promoter at its genomic locus (Fig. 3D). Thus, Rpn10 escapes degradation even though it is physically accessible to the proteasome's proteolytic machinery because it is constructed from a sequence that is not recognized by the proteasome.

Is the same mechanism used to ensure the stability of other proteasome subunits? Rpn3 is a structural subunit of the regulatory particle and, just as Rpn10, it contains a disordered C-terminal tail that appears to have the potential to reach the entrance to the translocation channel but remains stable (Fig. 4A). Thus, Rpn3 may also escape degradation because the amino acid sequence of the tail is not recognized by the proteasome. Indeed, HA-tagged Rpn3 was degraded by the proteasome when its C- terminal tail (amino acids 478-523) is replaced with an effective initiation sequence (Su9; Fig. 4B and C) without the need for ubiquitination (Supplementary Fig. 2A). HA- tagged Rpn3 with its native tail or with a tail that is not recognized by the proteasome (SP2; Fig. 4B and C) remain stable. The Rpn3 and Rpn10 derivatives were incorporated into the proteasome, at least to some extent, because they could be detected by western blotting after immunoprecipitation of proteasome particle through a FLAG-tag Rpn11 (Supplementary Fig. 3). Thus, Rpn3 too is protected from degradation by the amino acid sequence of its C-terminal tail.

*Unstructured tails of proteasome subunits are not efficient initiation sequences:*

As Rpn10 and Rpn3 are protected from degradation because the relevant part of their polypeptide chain is not recognized by the proteasome's translocation machinery, we asked whether other subunits are also constructed from sequences with similarly stealthy properties? To address this question, we fused the disordered regions from the C-termini of regulatory particle subunits that were at least 20 amino acids long to a model proteasome substrate and tested its degradation by the proteasome in cells. The model proteins were built on a yellow fluorescent protein (YFP) scaffold to simplify their detection in yeast and an N-terminal ubiquitin-like domain (UBL) of Rad23 to target to the proteasome. Rad23 is a non-stoichiometric proteasome subunit that associates with the regulatory particle through its UBL domain22. The UBL-YFP fusion proteins were not degraded in yeast unless a disordered region was also present in the proteins to allow the proteasome to initiate degradation (Fig. 5A)22. We attached the disordered regions derived from regulatory particle subunits to the C-terminus of the UBL-YFP fusion proteins and expressed these proteins together with a red reference protein from a CEN/ARS plasmid in yeast driven by TPI1 and PGK1 promoters, respectively. The yellow fluorescence of each cell, relative to its red fluorescence, determined by flow cytometry, served as a reliable measure of model substrate's proteasomal degradation22.

A UBL-YFP protein with a tail that is recognized by the proteasome was degraded so efficiently that YFP fluorescence was as low as background (Su9 in Fig. 5A). Degradation is due to the proteasome and is attenuated by proteasome inhibitor (Supplementary Figure 4A). In contrast, a UBL-YFP protein with a tail that is not recognized by the proteasome accumulates so that cells fluoresce many-fold over background (SP2 in Fig. 5A). The tails derived from the C-termini of regulatory particle subunits accumulated to more than 30-fold over background, suggesting that these polypeptide sequences avoid recognition by the proteasome (Fig. 5A). The constructs with the C-terminal disordered regions of Rpn3, Rpn10 and Rpn13 were the most stable proteins, and as stable as the test substrates without a tail (Fig. 5A). Interestingly, the C-terminal tail of Rpn2, and Rpn13, which does not seem to be able to reach the entrance pore, also escaped proteasome recognition. Moreover, even two copies of the disordered region of Rpn13 (or Rpn8, the shortest of the tails) escape recognition by the proteasome (Supplementary Fig. 4B). Thus, these proteins appear to be protected by two mechanisms: physical (the tails are too short to reach the channel) and chemical (even when the sequences can reach the entrance to the degradation channel, their amino acid sequence prevents recognition).

In addition to terminal-mediated degradation, the proteasome can initiate degradation from an internal region in a protein21,26. In fact, the proteasome subunit Rpn1 contains a large disordered loop (amino acids 626-735). To mimic a proteasome substrate with an internal disordered loop with both ends anchored in a folded domain, we attached the internal Rpn1 loop sequence to UBL-YFP and fused a BFP domain to the C-terminal end of the polypeptide. The proteasome was not able to initiate degradation at the Rpn1 loop or a negative control (SRR), even though a control sequence derived from a natural proteasome substrate (Spt2327 ) allowed degradation (Fig. 5B). In summary, all the disordered region of subunits in the proteasome cap tested here have sequences that resist degradation.

*Impact of degradation of essential proteasomal subunit Rpn3*

Finally, we asked whether the protective proteasome sequences we discovered are important physiologically. Rpn3 is a structural component of the proteasome complex and essential for proteasome function and thus cellular viability. Therefore, it may be necessary for Rpn3 to have a disordered sequence that is proteasome resistant *in vivo*. If so, expression of Rpn3 in which the degradation resistant sequence is replaced by a sequence that allows recognition should compromise the integrity of the proteasome and thus cell viability.

We tested this hypothesis by inserting either an efficient proteasome engagement sequence (HA-Rpn3ΔC-Su9) or an inefficient engagement sequence (HA- Rpn3ΔC-SP2) at the C-terminus of Rpn3 and predicted that the Su9 tail would cause loss of cell viability when the mutant subunit is not covered by a wildtype copy in the genome, whereas the SP2 tail would support viability. The constructs were carried on centromeric plasmids in cells with a heterozygous deletion of *RPN3* at the chromosomal locus (strain CMY 3749, Supplementary Table 1) and cell viability was determined by sporulation followed by tetrad analysis (see Methods). Deletion of *RPN3* was confirmed to be lethal, with the exception of a single viable spore, (see comments below). Viability was restored by a plasmid expressing HA-Rpn3 with the native C-terminus (strain CMY 3751) and we recovered 38 viable *rpn3Δ* spores carrying the plasmid (Table 3). Neither the empty vector (strain CMY 3750) nor the vector expressing HA-Rpn3ΔC-Su9 supported cell viability, although we again recovered a single viable *rpn3Δ* spore, this time carrying the HA-Rpn3ΔC-Su9 plasmid. In contrast, we recovered 20 viable *rpn3Δ* spores carrying the plasmid expressing HA-Rpn3ΔC-SP2. The unexpected observation of viable spores whose nominal genotype should not allow the accumulation of Rpn3 is probably the result of chromosome mis-segregation events. It is known that the proteasome is responsible for the degradation of chromosomal cohesins and therefore has a direct role in faithful chromosome segregation28. Therefore, proteasome impairment leads to higher rates of mis-segregation. In summary, we conclude that the native C-terminus of Rpn3 seems to have evolved to resist proteasomal degradation and ensure proteasome integrity and thus function.

# Discussion

The half-life of proteins can vary from minutes to days in cells. Regulatory proteins have short half-lives, whereas structural proteins and large multi-subunit protein complexes such as the proteasome have longer half-lives29,30. The assembly of large multi-subunit protein complexes is a significant challenge and requires the coordination of synthesis of subunits that at times may not fold into their native structures until late in the assembly process. The subunits have to be present at the correct stoichiometry, with unassembled supernumerary subunits cleared from the cell to avoid unwanted unbalanced activity or other toxic effects31,32. For this reason, the expression and assembly of proteasomal subunits is orchestrated, and the transcription of most subunits and the assembly chaperones is regulated by a common transcription factor, Rpn4 in yeast and Nrf1 in mammalian cells; activity of these transcription factors in turn is finely tuned33–36.

Proteasome subunits have noticeably long half-lives. Proteasome activity is tuned by post-translational modifications, proteasome autophagy (also referred to as proteophagy), and by the reversible

sequestration of assembled particles into granules37,38. During some forms of cellular stress such as glucose starvation and in aged non-dividing yeast cells, proteasome particles accumulate in the cytoplasm to form cytosolic granules called proteasome storage granules (PSGs) to escape autophagy39–42. The PSG formation appears to enhance resistance to genotoxic stress and upon glucose replenishment the granules dissolve rapidly and the proteasome particles become functional again41,43. Only under more extreme conditions such as sustained nitrogen starvation in *S. cerevisiae* and *Arabidopsis* or amino acid starvation in mammalian cells do proteasome particles undergo proteophagy, which leads to disassembly of the particles and proteolysis of the subunits into their amino acid components44–47.

The proteasome is a powerful protease with broad specificity, able to degrade effectively any protein, native or foreign. Most proteins are targeted for proteasomal degradation by ubiquitin tags and their primary role seems to be to localize the target protein to the proteasome. Ubiquitin-binding to the proteasome can stimulate its ATPase activity48, but localization to the proteasome is generally thought to be sufficient to induce protein degradation8,9. This raises the question as to how proteasome subunits, especially those located close to the entrance of the proteasome's substrate channel, such as the substrate receptors themselves, escape degradation.

Proteasome degradation requires the presence of a disordered region in a substrate to allow the proteasome to initiate degradation. Thus, it would in principle be possible to construct the proteasome from subunits that lack disordered regions. However, these intrinsically disordered regions often play important regulatory roles as the sites of post-translational modifications and interaction sites for binding partners49–51. Indeed, this is also the case for the proteasome15–17. Thus, the proteins that form an integral part of the proteasome complex are constantly exposed to the risk of being pulled into the degradative machinery and destroyed. Degradation of individual subunits would compromise the functional integrity of the proteasome and may even lead to disassembly if the particular subunit degraded plays a structural role. The proteasome is large and highly abundant so that synthesis of the subunits also sequestrates substantial resources in terms of energy and precursors52. Therefore, it is particularly expedient to prevent the churn of proteasome subunits.

Here we find that accidental self-disassembly by the proteasome is achieved by constructing disordered regions of subunits from amino acid sequences that are only poorly recognized by the proteasome. This strategy has been observed for other proteins, such as the ubiquitin conjugating enzyme Cdc34, which catalyzes its own ubiquitination on a long, disordered tail when overexpressed. The amino acid sequence of this tail is heavily biased and negatively charged and the proteasome is unable to initiate degradation on the tail53,54. Similarly, the UBL-UBA protein Rad23 binds to the proteasome through its UBL domain yet the proteasome fails to initiate degradation at any of the several disordered linkers present throughout Rad23, either because these linkers are too short to allow the proteasome to interact with them effectively, or because their amino acid sequence again resists proteasomal recognition53,55. Some subunits are placed far from the entrance to the degradation channel so that the proteasome seems physically prevented from reaching them and initiating degradation (e.g. Rpn2 and Rpn13). However, even in these

cases the tails are constructed from sequences that cannot be recognized by the proteasome (Fig 2A, B and 5A).

It is possible that some subunits contain disordered regions that are buried by an interaction with other subunits and thus not accessible to initiate degradation. In human cells, the C-terminal tail of Rpn10 might be masked by an interaction with the E3 ligase E6AP56 and the C-terminus of Rpn13 might be rendered inaccessible by its interaction with the deubiquitinating enzyme Uch37/ UCHL557–59. A similar mechanism was proposed to stabilize Mdy2. The Mdy2 contains a UBL domain flanked by a disordered region, but this region is masked by its binding partner Get4. Deletion of Get4 from the cell exposes the disordered region of Mdy2 that leads to its degradation60. Other possibilities are that ligand removal exposes an originally buried region that can target the protein for proteasomal degradation *in vitro*61 or, as is the case for Retinoblastoma protein (Rb), that calpain cleavage exposes a sequence that can be recognized by the proteasome62.

In summary, we find that proteasomal subunits have evolved to adopt several strategies described above to protect themselves from self-destruction. We have shown how the proteasome and likely similar other compartmentalized proteases escape proteasomal degradation by deploying mechanisms that prevent the protease from initiating degradation on its own subunits.

# Methods

*Modeling of unstructured tails in the 26S proteasome*

Complete atomic models of the proteasome were built using homology modeling software developed in-house. Models were constructed using the recent cryo-electron microscopy structures of the yeast 26S proteasome as starting points (PDB: 6FVU and 6FVT)18. A number of residues are unresolved at the N- and C-termini of many of the 47 protein chains. Missing residues were therefore built as follows. First, a putative conformation of the single internal loop region that was not resolved in the structures – that of residues 635 to 698 in Rpn1 – was added using the program Loopy63. Each tail was then constructed using libraries of two-residue fragments extracted from a non- redundant list of high-resolution protein structures available in the PDB (http://www.rcsb.org;19 and compiled by the PISCES server (on 16 May 2019)20. A total of 6791 protein chains solved at a resolution of 1.6 Å or better, with percent identities of 50 % or lower, and with R-factor values of 0.25 or better was downloaded. Libraries of conformations were then assembled for all 400 (20 x 20) possible two-residue fragments by extracting their coordinates from each of the 6791 protein structures. During the construction of these libraries, candidate two-residue fragments were eliminated from consideration if either residue had an assigned occupancy less than 1.0, if the two residues were not contiguous in sequence, if their Cα-Cα distances exceeded 4.0 Å, or if they included either the first or last residue in the protein chain. Finally, since the proteasome tails are considered to be predominantly unstructured, the coordinates of candidate two-residue fragments were only retained if both residues were determined to be adopting “coil” conformations. This was determined by first running the secondary structure analysis program dssp64 on each protein chain in the library;

dipeptides were then rejected if either of the two residues were given the following secondary structure assignments: H (helix), E (sheet), B (beta-bridge), or G (3,10-helix).

Having constructed libraries of conformations for each possible dipeptide, the N- terminal and C-terminal tails were built in stepwise fashion by progressively superimposing conformations sampled from the libraries. For example, to construct the C-terminal tail of Rpn5 (uniprot ID Q12250), whose last 5 residues are unresolved in the 6FVU pdb file, we proceed as follows. The sequence of the last 6 residues in the tail is HGLQAK with H440 being the last of the resolved residues in the chain. To build coordinates for the next residue in the chain, G441, a conformation of the two-residue fragment HG was randomly selected from its library, the backbone heavy (i.e. non- hydrogen) atoms of the H residue were superimposed on to those of H440, and the resulting coordinates of the accompanying G residue were accepted and added to the growing chain if: (a) none of its heavy atoms was within 1.75 Å of any previously placed heavy atom in the entire 26S proteasome, and (b) if the RMSD of the four superimposed backbone heavy atoms was <0.25 Å. If both conditions held, the conformation of G441 was accepted and the process was repeated with the aim of adding L442 by sampling conformations of the two-residue fragment GL and superimposing the backbone heavy atoms of the G residue as before. If, however, either condition was not met, a different conformation of the two-residue fragment HG was randomly selected and the procedure repeated. For each new residue to be added a maximum of 50 attempts was allowed; if no acceptable conformation could be found then all coordinates of the tail that had been constructed thus far were discarded and the process of growing the tail was begun again. In the unlikely event that the total number of failed attempts to build the tail exceeded 10,000 the tail was extended by one residue, by eliminating the last structured residue, and the process was repeated.

For all of the different constructs considered in this work a total of 100 different models of the entire 26S proteasome was constructed. In each individual model, the tails were built in a randomized order to ensures that there is not a systematic bias in some chains being allowed greater access to the proteasome interior through always being the first to be placed. Once constructed, the end-to-end length of each unstructured tail was recorded as the distance between the Cα atom of the last resolved residue at one end of the tail and the N- or C-terminal Cα atom at the other end of the tail. In addition, the distance between the Cα atom of the last resolved residue and a site roughly centered in the middle of the entrance to the ATPase ring was measured. Comparison of these two distances allows a rough estimate to be gained of the ease with which the unstructured tail might, in principle, be able to reach to the ATPase ring. We note that a more direct estimate could, in principle, be obtained by attempting to build conformations in which each tail of interest explicitly terminates at the entrance to the ATPase ring. While possible in principle, building such conformations in a way that does not incur some kind of steric clash with the remainder of the proteasome structure would be a very formidable undertaking.

*Expression constructs for proteasome subunits and other proteins*

The coding sequences of *S. cerevisiae* proteasomal subunits (Rpn3, Rpn10 and Rpn13) were cloned into CEN/ARS plasmid. Hemagglutinin (HA) tag was attached to the N- terminus by a linker (GS). The

unstructured C-terminus of Rpn3 and Rpn10, was replaced by the Su9 or SP2 sequence (Fig. 3B and 4B and Supplementary Table 2). The initiation regions were derived from subunit 9 of the $F_o$ component of the *N. crassa* ATPase (Su9) and a peptide region 2 of influenza A virus M2 protein (SP2)22. The internal initiation sequences were derived from Spt23 (645-695) and serine-rich region derived from herpes virus 1 ICP4 (SRR). HA-tagged proteasomal subunits and their derivatives were expressed from an inducible GAL1 promoter.

The fluorescence substrates were expressed from a constitutive triosephosphate isomerase (TPI1) promoter and a reference fluorescence protein (dsRed express 2) was expressed from a phosphoglycerate kinase 1(PGK1) promoter on a CEN/ARS plasmid22. The UBL domain of *S. cerevisiae* Rad23 was fused to the yellow fluorescent protein (YFP) by a linker sequence (VDGGSGGGS), and the C-terminal tails derived from the unstructured region of proteasomal subunits; Su9 and Sp2 sequences were attached to YFP by a short linker sequence (PR) (Supplementary Table 3)22.

*Yeast transformation and growth*

Expression plasmids were transformed into *S. cerevisiae* strains (Supplementary Table 1) using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) and selected on SD -Uracil agar plates supplemented with 2 % glucose.

The expression of N-terminally HA-tagged Rpn3, Rpn10, Rpn13 and their derivatives was induced by switching the growth media from SD -Uracil supplemented with 2 % raffinose to 2 % galactose and growing the cultures at 30 ˚C. For proteasome inhibition experiments, cells were treated with either DMSO or 100 μM MG132 (Calbiochem) for 2 h before harvesting.

*Western blot*

Exponentially growing cells were harvested by centrifugation at 8,000 g for 30 s at 4 °C and pre-treated with 2.0 M LiAc followed by 0.4 M NaOH for 5 min on ice. Cells were suspended in 100 μl of lysis buffer (0.1 M NaOH, 0.05 M EDTA, 2 % SDS, 2 % β-mercaptoethanol) supplemented with protease inhibitor cocktail (Millipore) and were incubated at 90 °C for 10 min. Cell lysates were neutralized by adding 4 M acetic acid, incubated at 90 °C for 10 min and then cleared by centrifugation for 10 min at 16,000 g65,66. The supernatant was resolved on SDS-PAGE and transferred to nitrocellulose membrane for Western blotting.

HA-tagged proteasomal subunit derivatives expressed under the control of GAL1 promoter were probed with anti-HA antibody (Roche Life Science). Proteasomal subunit (3xFLAG-Rpn10 and 3xFLAG-Rpn13) derivatives expressed from their native promoters were detected by a monoclonal anti-FLAG antibody (Cell Signaling). Endoplasmic reticulum membrane protein Scs2 served as the loading control and was detected by an anti-Scs2 polyclonal antibody (a gift from J. Brickner, Northwestern University). Alexa-680-labeled goat anti-rabbit antibody (Invitrogen) and Alexa-800-labeled goat anti- mouse antibody

(Rockland Immunochemicals) were used as secondary antibody. Protein amounts were measured by direct infrared fluorescence imaging (Odyssey LI- COR Biosciences).

*Flow cytometry*

Yeast cells were grown in SD -Uracil supplemented with 2 % glucose at 30 °C to early log phase. The fluorescence signals of RFP (dsRed express2) and YFP were measure using an LSR Fortessa (BD Biosciences). The data was analyzed using FlowJo software (BD Biosciences) to calculate median YFP over RFP fluorescence ratio of at least 10,000 cells from each population. Each assay was repeated at least three times.

*Yeast genetics*

To express *RPN10* and *RPN13* genes from their native promoters, coding regions of *RPN10* and *RPN13* genes in *pdr5Δ* cell (Supplementary Table 1) were replaced with N-terminal 3xFLAG-Rpn10 and 3xFLAG-Rpn13 with native or modified C-terminus using homologous recombination. The His3MX marker was used for positive selection. All the constructs were confirmed by Sanger sequencing.

*In vivo* functional tests of modified Rpn3 C-terminal tails were performed in diploid yeast with one copy of the *RPN3* gene deleted (Transomics technologies)67. For tetrad analysis, the diploid cells were sporulated and individual spores isolated by microdissection68. Individual spore genotypes were determined by replica-plating to the appropriate selective media.

# Declarations


## Acknowledgements

We acknowledge the Center for Biomedical Research Support at The University of Texas (CBRS) for providing access to research instrumentations. Matouschek lab members for their critical comments. We thank Dr. Brickner from Northwestern University for anti-Scs2 polyclonal antibody, and Dr. Patricia Clark from The University of Notre Dame for initiating the U54GM105816 funded protein translation research network (PTRN) (to AKSG, AM and AHE). This work was supported by grants R01GM124501from NIGMS and R21CA196456 NCI as well as grant F-1817 from the Welch foundation (to AM), and by grant R35GM122466 (to AHE).


## Authors contribution

AKSG, HY, AHE and AM conceived the study, designed the experiments, and interpreted data. AKSG and HY performed the experiments. AHE performed the homology modeling work and CY did Rpn3 dissection experiment. AKSG and AM wrote the manuscript with inputs from all the authors.

## Competing interests

AM is a paid consultant of Kymera Therapeutics. The other authors declare no competing interests.

# References


1. Schwartz, A. L. & Ciechanover, A. Targeting Proteins for Destruction by the Ubiquitin System: Implications for Human Pathobiology. *Annual Review of Pharmacology and Toxicology* **49**, 73–96 (2009).
2. Collins, G. A. & Goldberg, A. L. The Logic of the 26S *Cell* **169**, 792–806 (2017).
3. Yu, H. & Matouschek, A. Recognition of Client Proteins by the Proteasome. *Annual Review of Biophysics* **46**, 149–173 (2017).
4. Mettlen, M., Chen, P.-H., Srinivasan, S., Danuser, G. & Schmid, S. L. Annual Review of Biochemistry Regulation of Clathrin-Mediated Endocytosis. *Annual Review of BiochemistryBiochemistry* **86**, 637–657 (2018).
5. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. & Wolf, D. H. The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *Journal of Biological Chemistry* **272**, 25200–25209 (1997).
6. Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annual Review of Biochemistry* **78**, 477–513 (2009).
7. Komander, D. & Rape, M. The Ubiquitin Code. *Annual Review of Biochemistry* **81**, 203– 229 (2012).
8. Janse, D. M., Crosas, B., Finley, D. & Church, G. M. Localization to the proteasome is sufficient for degradation. *Journal of Biological Chemistry* **279**, 21415–21420 (2004).
9. Wilmington, S. R. & Matouschek, A. An inducible system for rapid degradation of specific cellular proteins using proteasome adaptors. *PLoS ONE* **11**, (2016).
10. Tomita, T. & Matouschek, A. Substrate selection by the proteasome through initiation regions. *Protein Science* **28**, 1222–1232 (2019).
11. Finley, D., Chen, X. & Walters, K. J. Gates, Channels, and Switches: Elements of the Proteasome Machine. *Trends in Biochemical Sciences* **41**, 77–93 (2016).
12. Tanaka, K. & Ichihara, A. Half-life of proteasomes (multiprotease complexes) in rat *Biochemical and Biophysical Research Communications* **159**, 1309–1315 (1989).
13. Schwanhüusser, B. *et al.* Global quantification of mammalian gene expression *Nature* **473**, 337–342 (2011).
14. Tomita, T. *et al.* Specific Modification of Aged Proteasomes Revealed by Tag- Exchangeable Knock-In Mice. *Molecular and Cellular Biology* **39**, (2018).
15. Hirano, H., Kimura, Y. & Kimura, A. Biological significance of co- and post-translational modifications of the yeast 26S proteasome. *Journal of Proteomics* **134**, 37–46 (2016).
16. Kors, S., Geijtenbeek, K., Reits, E. & Schipper-Krom, S. Regulation of Proteasome Activity by (Post-)transcriptional Mechanisms. *Frontiers in Molecular Biosciences* **6**, 48 (2019).

17. Finley, D. & Prado, M. A. The proteasome and its network: Engineering for *Cold Spring Harbor Perspectives in Biology* **12**, (2020).

18. Eisele, M. R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell Reports* **24**, 1301–1315 (2018).

19. Berman, H. M. The Protein Data Bank / Biopython. *Presentation* **28**, 235–242 (2000).

20. Wang, G. & Dunbrack, R. L. PISCES: A protein sequence culling server. *Bioinformatics* **19**, 1589–1591 (2003).

21. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nature Structural and Molecular Biology* **11**, 830–837 (2004).

22. Yu, H. *et al.* Conserved sequence preferences contribute to substrate recognition by the proteasome. *Journal of Biological Chemistry* **291**, 14526–14539 (2016).

23. Filigheddu, N. *et al.* Santiago M. Di Pietro,* Juan M. Falco ´n-Pe ´rez,* Danie `le Tenza,† Subba R.G. Setty,‡ Michael S. Marks,‡ Grac ¸a Raposo,† and Esteban C. Dell'Angelica*. *Molecular biology of the cell* **18**, 986–994 (2007).

24. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).

25. Bachmair, A. & Varshavsky, A. The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032 (1989).

26. Liu, C. W., Corboy, M. J., DeMartino, G. & Thomas, P. J. Endoproteolytic activity of the proteasome. *Science* **299**, 408–411 (2003).

27. Piwko, W. & Jentsch, S. Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site. *Nature Structural and Molecular Biology* **13**, 691–697 (2006).

28. Rao, H., Uhlmann, F., Nasmyth, K. & Varshavsky, A. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410**, 955–959 (2001).

29. Price, J. C., Guan, S., Burlingame, A., Prusiner, S. B. & Ghaemmaghami, S. Analysis of proteome dynamics in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14508–14513 (2010).

30. Cambridge, S. B. *et al.* Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *Journal of Proteome Research* **10**, 5275–5284 (2011).

31. Peters, L. Z. *et al.* The Protein Quality Control Machinery Regulates Its Misassembled Proteasome Subunits. *PLoS Genetics* **11**, (2015).

32. Sung, M. K. *et al.* A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. *eLife* **5**, (2016).

33. Mannhaupt, G., Schnall, R., Karpov, V., Vetter, I. & Feldmann, Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Letters* **450**, 27–34 (1999).
34. Saeki, Y. & Tanaka, K. Assembly and function of the proteasome. *Methods in Molecular Biology* **832**, 315–337 (2012).
35. Shirozu, R., Yashiroda, H. & Murata, S. Identification of minimum Rpn4-responsive elements in genes related to proteasome functions. *FEBS Letters* **589**, 933–940 (2015).
36. Radhakrishnan, S. K. *et al.* Transcription Factor Nrf1 Mediates the Proteasome Recovery Pathway after Proteasome Inhibition in Mammalian Cells. *Molecular Cell* **38**, 17–28 (2010).
37. Martin-Perez, M. & Villén, J. Determinants and Regulation of Protein Turnover in *Cell Systems* **5**, 283-294.e5 (2017).
38. Guo, X. *et al.* Site-specific proteasome phosphorylation controls cell proliferation and tumorigenesis. *Nature Cell Biology* **18**, 202–212 (2016).
39. Yedidi, R. S., Fatehi, A. K. & Enenkel, C. Proteasome dynamics between proliferation and quiescence stages of Saccharomyces cerevisiae. *Critical Reviews in Biochemistry and Molecular Biology* **51**, 497–512 (2016).
40. Lee, Y., Cheng, K. Y., Chao, J. C. & Leu, J. Y. Differentiated cytoplasmic granule formation in quiescent and non-quiescent cells upon chronological aging. *Microbial Cell* **3**, 109–119 (2016).
41. Laporte, D., Salin, B., Daignan-Fornier, B. & Sagot, I. Reversible cytoplasmic localization of the proteasome in quiescent yeast cells. *Journal of Cell Biology* **181**, 737–745 (2008).
42. Gu, C. *et al.* Ubiquitin orchestrates proteasome dynamics between proliferation and quiescence in yeast. *Molecular Biology of the Cell* **28**, 2479–2491 (2017).
43. Marshall, S. & Vierstra, R. D. Proteasome storage granules protect proteasomes from autophagic degradation upon carbon starvation. *eLife* **7**, (2018).
44. Cohen-Kaplan, V. *et al.* p62- and ubiquitin-dependent stress-induced autophagy of the mammalian 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E7490–E7499 (2016).
45. Marshall, S., Li, F., Gemperline, D. C., Book, A. J. & Vierstra, R. D. Autophagic Degradation of the 26S Proteasome Is Mediated by the Dual ATG8/Ubiquitin Receptor RPN10 in Arabidopsis. *Molecular Cell* **58**, 1053–1066 (2015).
46. Waite, K. A., De-La Mota-Peynado, A., Vontz, G. & Roelofs, J. Starvation induces proteasome autophagy with different pathways for core and regulatory particles. *Journal of Biological Chemistry* **291**, 3239–3253 (2016).
47. Marshall, S., McLoughlin, F. & Vierstra, R. D. Autophagic Turnover of Inactive 26S Proteasomes in Yeast Is Directed by the Ubiquitin Receptor Cue5 and the Hsp42 Chaperone. *Cell Reports* **16**, 1717–1732 (2016).

48. Peth, A., Kukushkin, N., Bossé, M. & Goldberg, A. L. Ubiquitinated proteins activate the proteasomal ATPases by binding to Usp14 or Uch37 homologs. *Journal of Biological Chemistry* **288**, 7781–7790 (2013).
49. Wright, P. E. & Dyson, J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **16**, 18–29 (2015).
50. Darling, A. L. & Uversky, V. N. Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Frontiers in Genetics* **9**, (2018).
51. Deiana, A., Forcelloni, S., Porrello, A. & Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS ONE* **14**, (2019).
52. Suraweera, A., Münch, , Hanssum, A. & Bertolotti, A. Failure of amino acid homeostasis causes cell death following proteasome inhibition. *Molecular Cell* **48**, 242– 253 (2012).
53. Fishbain, S. *et al.* Sequence composition of disordered regions fine-tunes protein half-life. *Nature Structural and Molecular Biology* **22**, 214–221 (2015).
54. Goebl, M. G., Goetsch, L. & Byers, B. The Ubc3 (Cdc34) ubiquitin-conjugating enzyme is ubiquitinated and phosphorylated in vivo. *Molecular and Cellular Biology* **14**, 3022–3029 (1994).
55. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S. & Matouschek, A. Rad23 escapes degradation because it lacks a proteasome initiation region. *Nature Communications* **2**, 192–199 (2011).
56. Buel, G. R. *et al.* Structure of E3 ligase E6AP with a proteasome-binding site provided by substrate receptor hRpn10. *Nature Communications* **11**, 1–15 (2020).
57. Qiu, X. B. *et al.* hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, *EMBO Journal* **25**, 5742–5753 (2006).
58. Hamazaki, J. *et al.* A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO Journal* **25**, 4524–4536 (2006).
59. Yao, T. *et al.* Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nature Cell Biology* **8**, 994–1002 (2006).
60. Yu, H., Kago, G., Yellman, C. M. & Matouschek, A. Ubiquitin-like domains can target to the proteasome but proteolysis requires a disordered region. *The EMBO Journal* **35**, 1522–1536 (2016).
61. Gautam, A. K., Balakrishnan, S. & Venkatraman, P. Direct ubiquitin independent recognition and degradation of a folded protein by the eukaryotic proteasomes-origin of intrinsic degradation signals. *PLoS ONE* **7**, (2012).
62. Tomita, T., Huibregtse, M. & Matouschek, A. A masked initiation region in retinoblastoma protein regulates its proteasomal degradation. *Nature Communications* **11**, 1–8 (2020).
63. Xiang, Z., Soto, C. S. & Honig, B. Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7432–7437 (2002).

64. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical feature *Biopolymers* **22**, 2577–2637 (1983).
65. von der Haar, T. Optimized protein extraction for quantitative proteomics of yeasts. *PLoS ONE* **2**, (2007).
66. Zhang, T. *et al.* An improved method for whole protein extraction from yeast Saccharomyces cerevisiae. *Yeast* **28**, 795–798 (2011).
67. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
68. Lingbeek, M. E., Jacobs, J. J. L. & Van Lohuizen, M. The T-box repressors TBX2 and TBX3 specifically regulate the tumor suppressor gene p14ARF via a variant T-site in the initiator. *Journal of Biological Chemistry* **277**, 26120–26127 (2002).

# Tables

*Table 1. Unstructured regions in proteasome subunits*

| Subunit | no of residues | distance from translocation channel | maximum tail length | ratio |
|---|---|---|---|---|
| Rpn1 | 23 | 76.06 | 53.93 | 0.71 |
| Rpn2 | 20 | 122.64 | 49.91 | 0.41 |
| Rpn3 | 31 | 71.90 | 57.50 | 0.80 |
| Rpn8 | 34 | 85.68 | 75.72 | 0.88 |
| Rpn10 | 71 | 124.41 | 128.70 | 1.03 |
| Rpn13 | 23 | 135.23 | 49.37 | 0.37 |

Residues: only terminal disordered residues are shown

Tail length (Å): maximum length recorded between Cα atom of the last resolved residue of the proteasomal subunit (PDB 6FVT) and the C-terminal Cα atom at the other end of the modeled unstructured peptide.

Distance from ATPase channel (Å): The distance between the Cα atom of the last resolved residue and a site roughly centered in the middle of the entrance to the translocation channel.

Ratio: likelihood of a disordered tail reaching the translocation channel, ratio of tail length and distance from the channel.

*Table 2. Degradation of Rpn3 is lethal for cells*

Test of *rpn3ΔC-tail* constructs for complementation of *rpn3Δ* lethality.

| strain | relevant genotype | # tetrads dissected | # viable *rpn3Δ spores* |
|---|---|---|---|
| CMY 3749 | *RPN3/rpn3Δ* | 39 | 1 |
| CMY 3750 | *RPN3/rpn3Δ* + empty vector | 39 | 0 |
| CMY 3751 | *RPN3/rpn3Δ* + *RPN3* vector | 39 | 38 |
| CMY 3752 | *RPN3/rpn3Δ* + *rpn3ΔC-SP2* vector | 52 | 20 |
| CMY 3753 | *RPN3/rpn3Δ* + *rpn3ΔC-Su9* vector | 52 | 1 |

# Figures

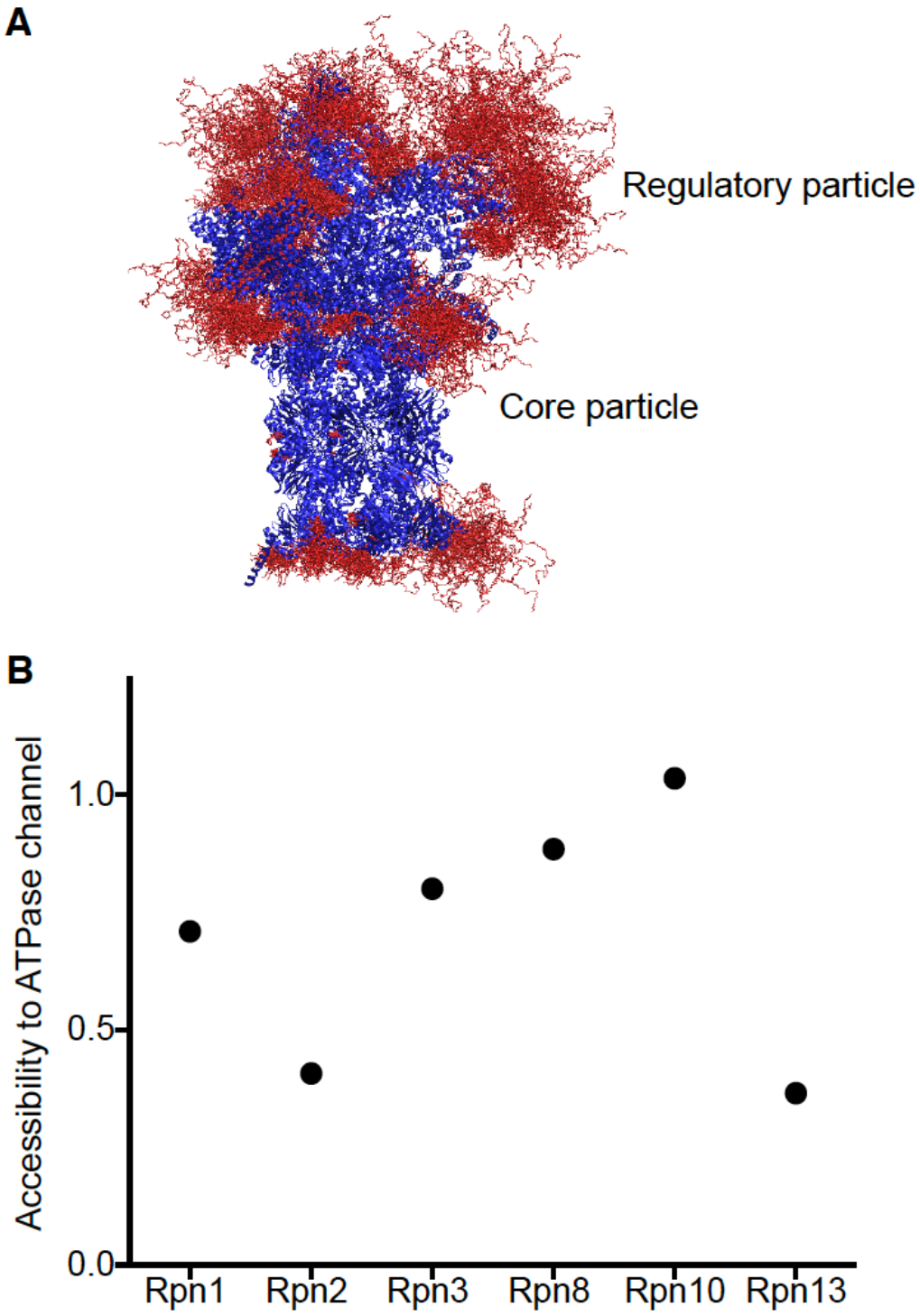


**Figure 1**

Proteasome subunits contain disordered regions that potentially reach the translocation channel (A) Molecular modeling of unresolved regions of proteasomal subunits (PDB: 6FVU); for each unresolved tail, one hundred different conformations are shown in red; all resolved residues are shown in blue. (B)

Accessibility of C-terminal disordered tails of Rpn1, Rpn2, Rpn3, Rpn8, Rpn10, and Rpn13 (also see Table 1) to the translocation channel was calculated by taking the ratio of length of disordered regions (Å) as calculated by molecular modeling and the distance of disorder region from the translocation channel (Å).

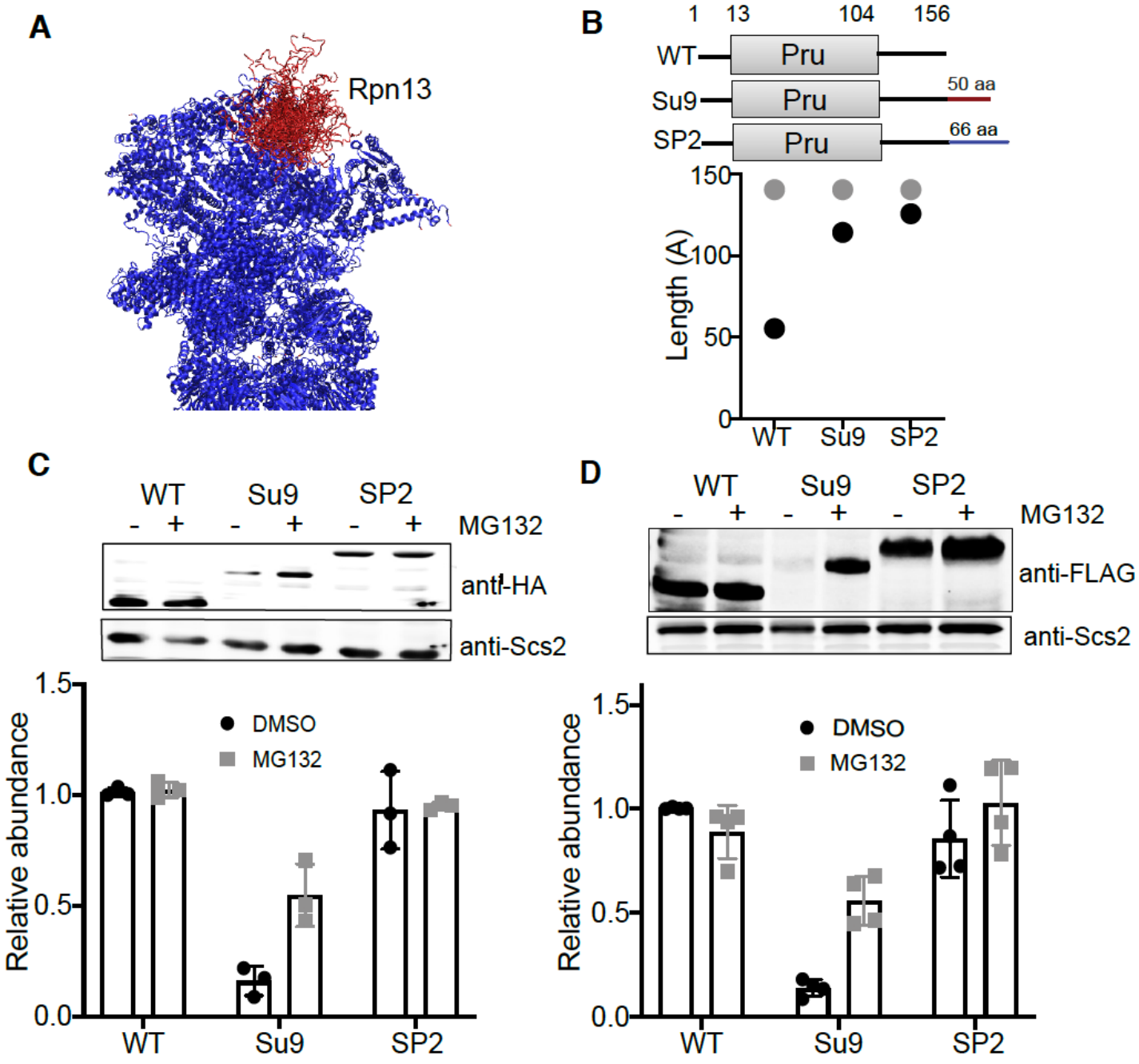


**Figure 2**

Rpn13 is protected as its unstructured tail is not accessible to the translocation channel (A) Molecular modeling of the proteasome (PDB: 6FVU). One hundred different conformations of the C-terminal disordered region of Rpn13 are shown in red; all other residues are shown in blue. (B) Schematic representation of Rpn13 and its derivatives. The C-terminal tail of Rpn13 was extended by attaching Su9 and SP2 sequences (also see Methods and Supplementary Table 2). The length of the C-terminal

disordered region of WT, Su9 and SP2 fusion of Rpn13 (Å) was calculated by molecular modeling (black circles) and compared with the distance (Å) of disorder tail from the translocation channel (grey circles). Pru: Pleckstrin-like receptor for ubiquitin domain. (C) The expression of HA-Rpn13 and its derivatives was induced by switching the cells to galactose supplemented media, cells were treated with either DMSO or 100 μM MG132 for 2 hours. In vivo protein abundance was measured by immunoblotting for HA and direct infrared fluorescence imaging. Scs2 serves as a loading control. Relative abundance was calculated by first normalizing band intensity with loading control, and with DMSO treated Rpn13 with native C-terminus. Error bars indicate mean ± standard deviations (S.D.). (D) The RPN13 genomic copy in pdr5Δ cell was replaced by sequence encoding 3xFLAG-Rpn13 with native or extended tail (Su9 or SP2). The cells were treated with either DMSO or 100 μM MG132 for 2 hours, and the protein level was measured by immunoblotting for FLAG. Scs2 serves as a loading control. Relative abundance was calculated by first normalizing band intensity with loading control, and with DMSO treated Rpn13 with native C-terminus. Error bars indicate mean ± S.D.

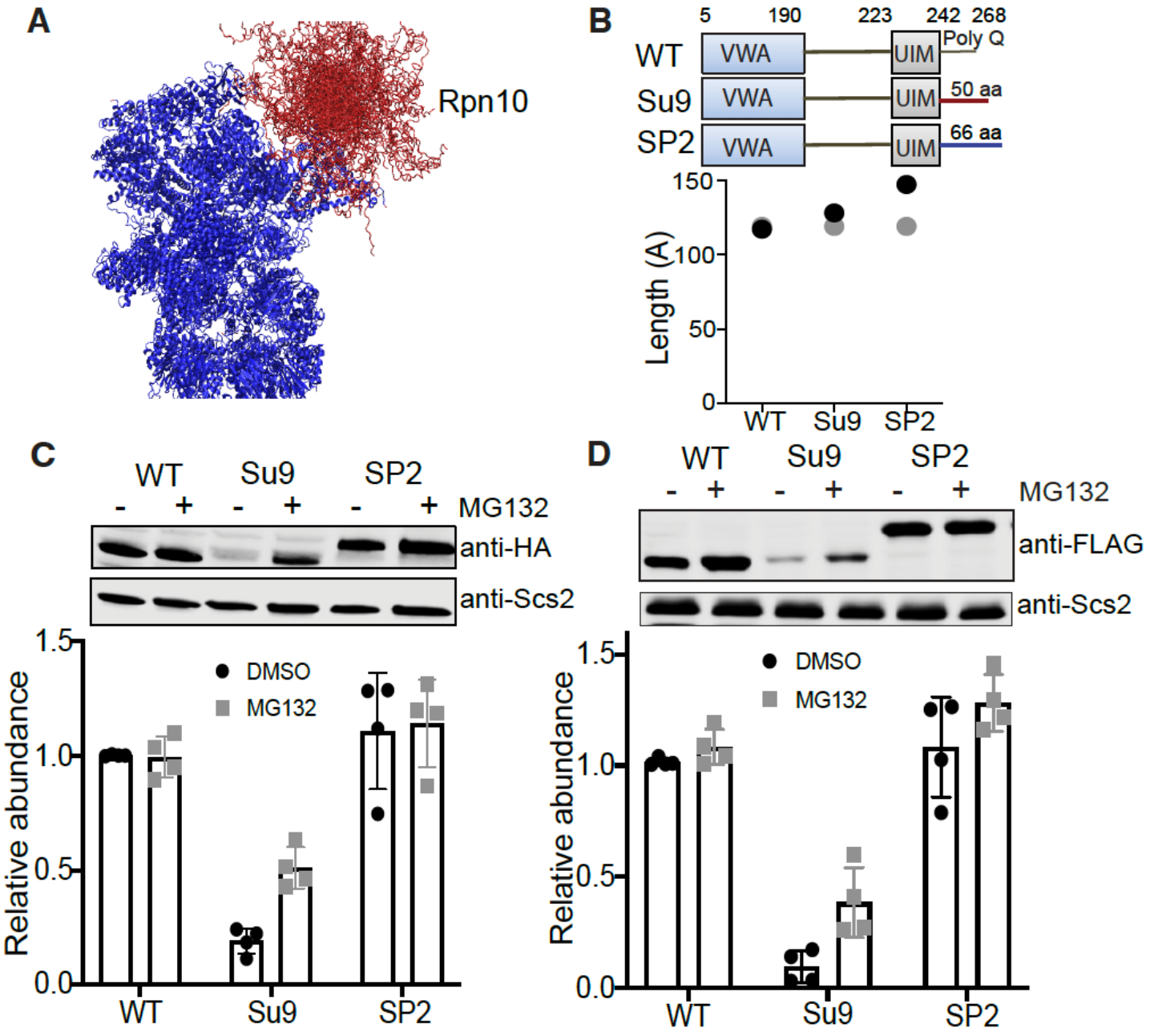


**Figure 3**

The sequence property of Rpn10's unstructured region protect itself from self- destruction (A) Molecular modeling of the proteasome (PDB: 6FVU). One hundred different conformations of the C-terminal disordered region of Rpn10 are shown in red; all other residues are shown in blue. (B) Schematic representation of Rpn10 and its derivatives. The C-terminus tail of Rpn10 is rich in glutamine amino acid (poly Q). The C-terminal 243-268 amino acids were replaced by either Su9 or SP2 (also see Methods and Supplementary Table 2). The length of C-terminal disordered region of WT, Su9 and SP2 fusion of Rpn10 (Å) was calculated by molecular modeling (black circles) and the distance (Å) of disorder tail from the

translocation channel (grey circles) was compared. VWA: von Willebrand factor type A domain and UIM: Ubiquitin Interacting Motif, (C) The expression of HA-Rpn10 and its derivatives was induced by switching the cells to galactose supplemented media, cells were treated with either DMSO or 100 μM MG132 for 2 hours. Protein abundance was measured by immunoblotting for HA and direct infrared fluorescence imaging. Scs2 serves as a loading control. Relative abundance was calculated by first normalizing band intensity with lading control, and with DMSO treated Rpn10 with native C-terminus. Error bars indicate mean ± S.D. (D) The RPN10 genomic copy in pdr5Δ cell was replaced by sequence encoding N-terminally 3xFLAG-tagged Rpn10 with native C-terminus or C-terminus (243-268) was replaced by Su9 or SP2. Cells were treated with either DMSO or 100 μM MG132 for 2 hours and the steady-state protein level was measured by immunoblotting for FLAG, Scs2 acts as a loading control. Relative abundance was calculated by first normalizing band intensity with loading control, and with DMSO treated Rpn10 with native C- terminus. Error bars indicate mean ± S.D.

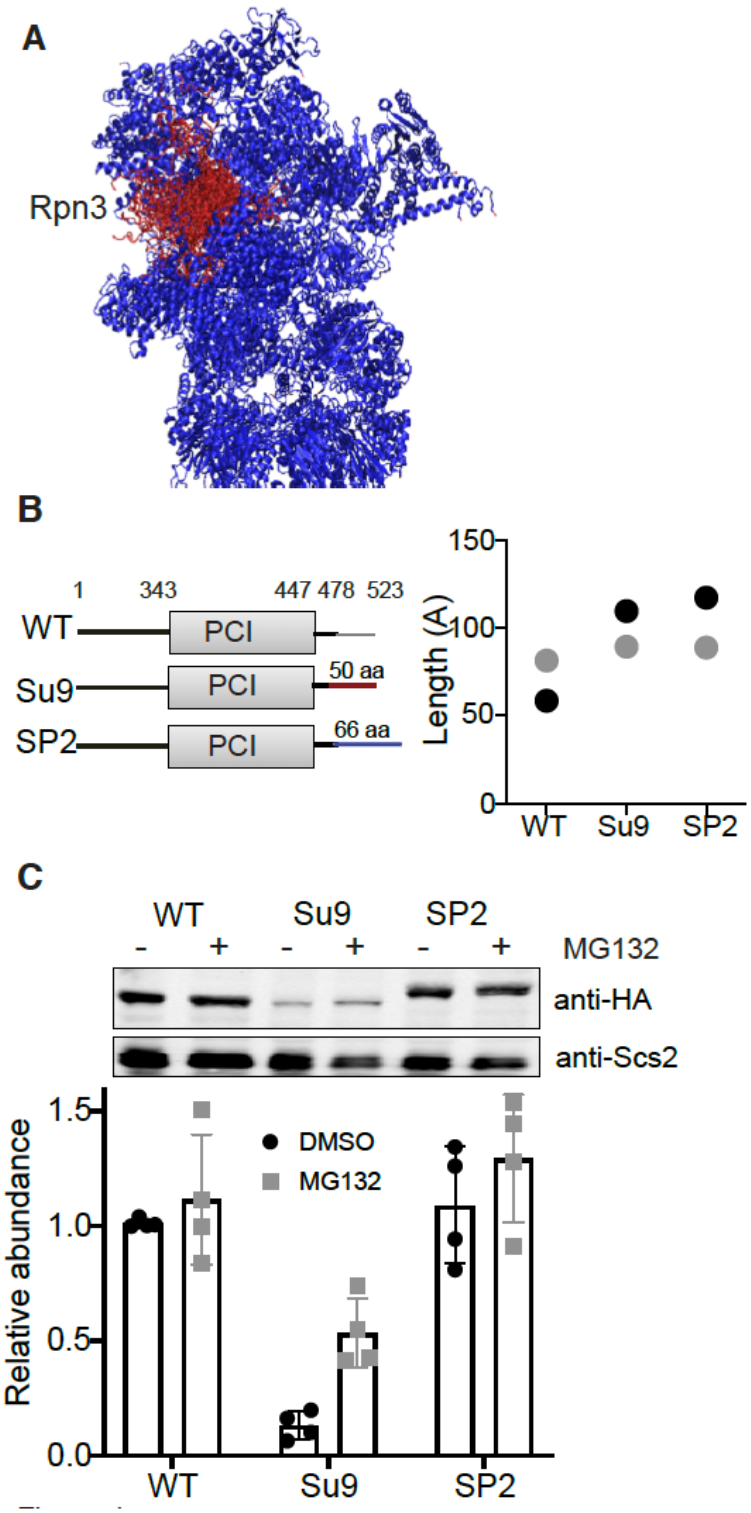


**Figure 4**

Rpn3 is protected by its amino acid sequence properties (A) Molecular modeling of the proteasome (PDB: 6FVU). One hundred different conformations of the C-terminal disordered region of Rpn3 are shown in red; all other residues are in blue (B) Schematic representation of Rpn3 and its derivatives. The C-terminus of Rpn3 (478- 523) was replaced by either Su9 or SP2 (see Methods and Supplementary Table 2). The length of the C-terminal disordered region of WT, Su9 and SP2 fusion of Rpn3 (Å) was calculated by

molecular modeling (black circles) and the distance (Å) of disorder tail from the translocation channel (grey circles) was compared. PCI: Proteasome, COP9, Initiation factor 3 domain. (C) The expression of HA-Rpn3 and its derivatives was induced by switching the cells to galactose supplemented media, cells were treated with either DMSO or 100 μM MG132 for 2 hours. Protein abundance was measured by immunoblotting for HA and direct infrared fluorescence imaging. Scs2 acts as a loading control. Relative abundance was calculated by first normalizing band intensity with loading control, and with DMSO treated Rpn3 with native C-terminus. Error bars indicate mean ± S.D.

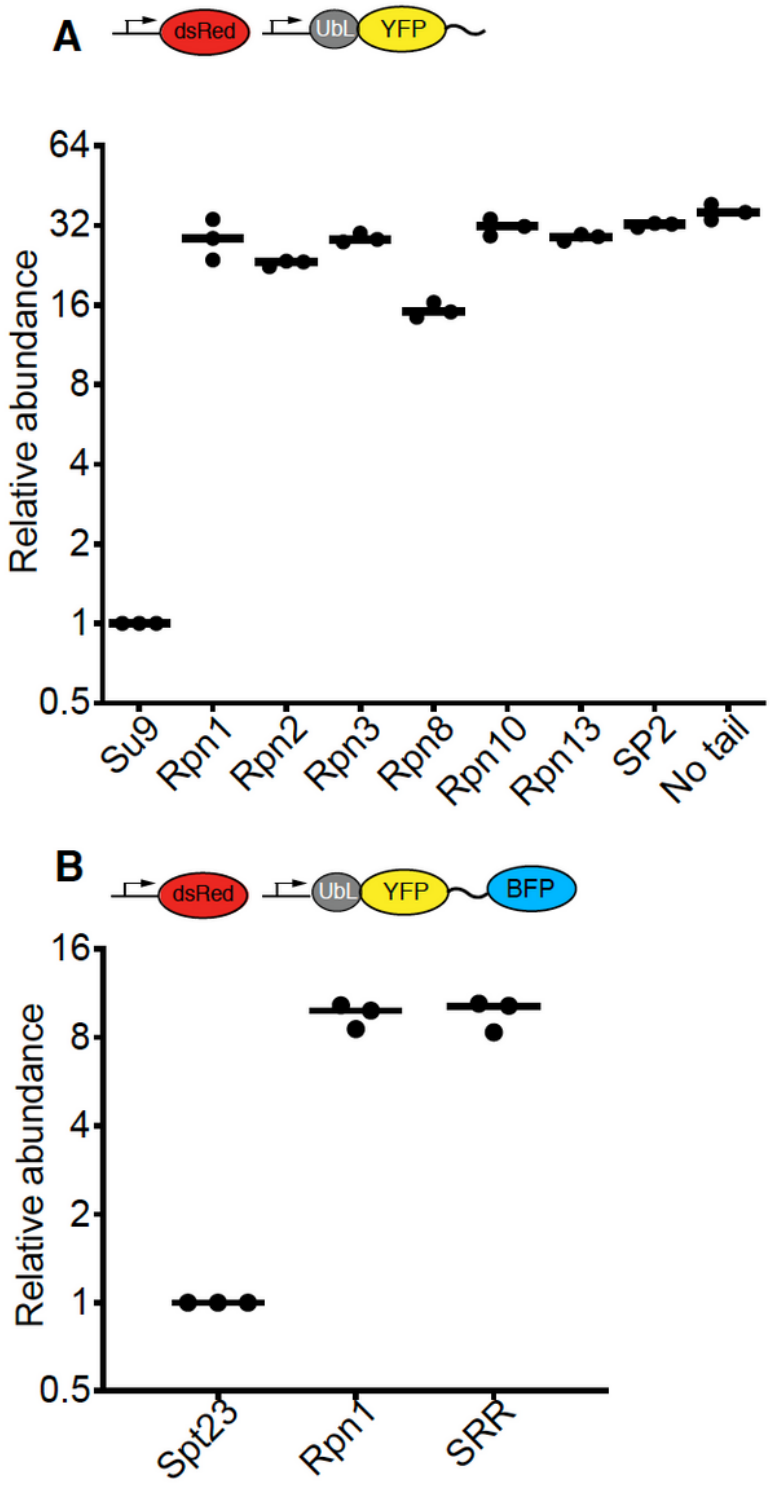

**Figure 5**

The unstructured tails of proteasome subunits are not efficient initiation sequences (A) Unstructured C-terminal tail of proteasomal subunits Rpn1 (963-993) Rpn2 (907- 945), Rpn3 (479-523), Rpn8 (309-338), Rpn10 (189-268) and Rpn13 (132-156) were fused to C-terminus of model protein UBL-YFP. The dsRed served as an internal control. The Su9 is an efficient initiation region while SP2 is inefficient initiation region. Median YFP over RFP (dsRed) florescent intensity from at least 10,000 cells was used to measure in vivo relative abundance of fusion proteins. Error bars indicate mean ± S.D. (B) Internal unstructured regions of Rpn1 (626-735), efficient internal initiation from Spt23 (645-695), and inefficient initiation sequence SRR were fused between UBL-YFP and BFP. The dsRed served as an internal control. Median YFP over RFP (dsRed) florescent intensity from at least 10,000 cells was used to measure in vivo relative abundance of fusion proteins. Error bars indicate mean ± S.D.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SinghGautametalSupplementarydocument.pdf