Title
Redefining HbYX-dependent activation of the human 20S proteasome

Permalink
https://escholarship.org/uc/item/8rq9g93k

Author
Opoku-Nsiah, Kwadwo Appiah

Publication Date
2020

Peer reviewed|Thesis/dissertation

Redefining HbYX-dependent activation of the human 20S proteasome

by
Kwadwo Appiah Opoku-Nsiah

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Jason Gestwicki   Jason Gestwicki
4909848DBB404E5...   Chair

DocuSigned by:

Michelle Arkin   Michelle Arkin
DocuSigned by: ...34...

Jack Taunton   Jack Taunton
0FD9D71201FC43C...

_____

_____
                                                    Committee Members

Dedicated to Junior and the Black narrative

Acknowledgements

I would first like to thank my advisor Jason Gestwicki for his guidance, encouragement and unwavering support in getting me to this point in my graduate study. Thank you for demonstrating unmerited grace as I managed through the challenges of science research during a particularly challenging period in my life.

Thank you to Andrej Sali and Peter Cimermančič for contacting Jason with such a bold scientific endeavor that initiated my scientific journey. Thank you to the many collaborators, particularly Andres de la Peña and Gabe Lander, that enabled our better understanding of this system. And then thank you to my thesis committee and the Institute for Neurodegenerative Diseases for supporting my scientific development along the way.

It has been a privilege to be a part of the Gestwicki Gang. Thank you to all the current and former members for being thoughtful scientists, amazing advocators, genuine people and lifelong colleagues and friends. This sentiment truly extends to many others in the UCSF community, for whom I am extremely grateful to have as a part of my community. Thank you for being a beacon of light throughout the darkest hours.

I could not have completed this work without the love and support of my family and my homies. Thank you for shaping me into the person I am today. To my parents Yaw and Afua Opoku-Nsiah, my siblings Cynthia, Kwaku, Atta Yaw and Adwoa, my cousins, Uncles and Aunties - *Meda wo ase*.

Acknowledgement of previously published material

Chapters 1 and 2 are reprints of the material as it appears in:

Opoku-Nsiah, K.A. and Gestwicki J.E. Aim for the Core: Suitability of the ubiquitin-independent 20S proteasome as a drug target in neurodegeneration. **Trans. Res.** 198, 48-57 (2018)

Opoku-Nsiah, K.A., de la Pena A.H., Williams S.K., Chopra N, Sali A, Lander G.C. and Gestwicki J.E. The YΦ motif defines the structure-activity relationships of human 20S proteasome activators. *Manuscript under review.*

Redefining HbYX-dependent activation of the human 20S proteasome

by

Kwadwo Appiah Opoku-Nsiah

Abstract

Cells must routinely recycle their proteins in order to maintain protein homeostasis and health. The bulk of protein turnover is facilitated by the proteasome and locally restricted within the center of the barrel-shaped 20S core. In order to be degraded, potential substrates must enter the core through an opening at the center of this complex. The pore is gated by N-terminal extensions that limit substrate entry and can be opened by native activators, which functionally regulate protein degradation. Gate-opening occurs through HbYX motifs at the extreme C-termini of proteasome activators. To date, characterization of the HbYX mechanism has been limited to protein-protein interactions of archaeal proteasome complexes. We have uncovered unexpected differences in the residue specificity of this motif for activating the human 20S proteasome. These sequence requirements are tuned by valency, which was generally corroborated by the C-termini of natural activators, providing insight into identifying new binding partners of the eukaryotic 20S proteasome. This work has established an alternative model to the HbYX mechanism for eukaryotic proteasome complexes. Finally, the YΦ motif offers an updated template for the design of pharmacological activators, critical protein-protein interface will deepen our understanding of proteasome biology and its implication in disease.

Table of Contents

List of Figures

List of Tables

Chapter 1

Roles of the ubiquitin-independent 20S proteasome in neurodegeneration and its suitability as a drug target

Kwadwo A. Opoku-Nsiah and Jason E. Gestwicki

Abstract

Neurodegenerative diseases are a class of age-associated proteinopathies characterized by the accumulation of misfolded and/or aggregation-prone proteins. This imbalance has been attributed, in part, to an age-dependent decay in the capacity of protein turnover. Most proteins are degraded by the ubiquitin-proteasome system (UPS), which is composed of ubiquitin ligases and regulatory particles, such as the 19S, that deliver cargo to the proteolytically active 20S proteasome core. However, a subset of clients, especially intrinsically disordered proteins (IDPs), are also removed by the action of the ubiquitin-independent proteasome system (UIPS). What are the specific contributions of the UPS and UIPS in the context of neurodegeneration? Here, we explore how age-associated changes in the relative contribution of the UPS and UIPS, combined with the IDP-like structure of many neurodegenerative disease-associated proteins, might contribute. Strikingly, the core 20S proteasome (20S) has been shown to predominate in older neurons and to preferentially act on relevant substrates, such as synuclein and tau. Moreover, pharmacological activation of the 20S has been shown to accelerate removal of aggregation-prone proteins in some models. Together, these recent studies are turning attention to the 20S proteasome and the UIPS as potential therapeutic targets in neurodegeneration.

Introduction to the Proteasome

The proteasome is a central protein degradation machine in eukaryotes[1]. Through hydrolysis activities, it removes damaged proteins and ensures the delivery of amino acids to support ongoing biosynthesis. In addition, the proteasome has been co-opted for

2

more specialized tasks in regulating the cell cycle, differentiation, the inflammatory response, antigen presentation and apoptosis[2,3]. To enable these functions, the proteasome makes up a staggering 1 to 2% of the entire proteome in healthy cells. However, a decline in proteasome activity has been broadly implicated in ageing and age-associated diseases, including neurodegeneration. Presumably, this decline contributes to a catastrophic imbalance in proteostasis and accumulation of damaged and/or misfolded proteins. In this review, we explore the structure-function of the proteasome and its implications in the onset and progression of neurodegenerative disease. In addition, we focus on emerging therapeutic opportunities through pharmacological activation of this degradation machine.

The 20S proteasome (20S) is a barrel-shaped complex comprised of four heptameric rings: two stacked β-rings that are sandwiched by two α-rings (**Figure 1.1A**). Three of the seven subunits (β1, β2, & β5) that make up the β-ring are proteases that hydrolyze peptide bonds of substrates. These active sites are sequestered in the interior of the 20S chamber, such that substrates must first traverse through the exterior α-rings. In its closed state, the α-rings have a narrow pore that occludes the entry of most proteins[4]. Thus, one key to understanding proteasome regulation is to learn how substrates are granted access to the proteolytic chamber. Substrates are targeted to the proteasome through two major pathways, the ubiquitin-proteasome system (UPS) and the ubiquitin-independent proteasome system (UIPS). Proteasome activators (PA), which are predominantly multi-protein complexes, help facilitate degradation by the 20S. There are many types of PAs and the specific one that is bound determines whether that 20S proteasome is coupled to the UPS or UIPS (**Figure 1.1A**). However, most of the PAs

share a conserved tripeptide sequence, the HbYX (Hydrophobic-TYRosine-unspecified residue 'X'), at their C-termini that interacts with pockets in the α-rings of the 20S to allosterically open the pore[5].

Ubiquitin-proteasome system (UPS). Proteasomal degradation by the UPS first requires the conjugation of multiple ubiquitin (Ub) proteins onto the substrate, generating the poly-Ub signal that designates it as a substrate of the proteasome. Recent work has shown that conjugation of two or more polyUb chains is needed on the tagged substrate to efficiently interact with the UPS machine[6]. Thus, regulation of this pathway by the activity of the E1, E2 and E3 Ub ligases is a critical component of its function[7], but will not be described in detail here. The canonical regulatory particle of the UPS is PA700 (or 19S), which is a 700 kDa proteasome activator complex that associates with the 20S to create the 26S proteasome (26S)[8]. PA700 is comprised of a 'base' and a 'lid'. The lid contains subunits that bind to polyUb chains, as well as deubiquitinating enzymes (DUBs) that regulate association with the particle. The base contains the HbYX motifs that interact with the α-rings, and ATPases that unfold the substrate so that it can access the proteolytic chamber[9]. Recent reviews provide additional information about the structure of the 26S and its biological fuction[6].

Ub-independent-proteasome system (UIPS). Ub-independent degradation is coordinated by the 20S and may be amplified with UIPS-specific PAs, including PA200 and the heptameric PA28[10]. PA200 is a monomeric protein that uses a C-terminal HbYX motif to bind to and activate the 20S. PA28 is composed of multiple, different subunits (alpha, beta, and gamma) and it relies on an alternative (e.g. non-HbYX) motif for association with the 20S[11,12]. The UIPS-specific PAs typically lack the unfolding activity

of PA700; rather, they open the α-ring gate through a binding-induced conformational change and increase the flux of suitable substrates into the proteolytic chamber[13]. As discussed below, this mechanism restricts UIPS substrates to unfolded proteins that can fit into the channel without an active unfoldase. However, PA700-bound 26S has an open α-ring gate too and is thus capable of facilitating ubiquitin-independent substrate turnover[14]. The relative contributions of the 26S in the UPS and UIPS pathways remain unclear (**Figure 1.1B**); and, for simplicity, we will only include the contributions of the 26S to UIPS in passing in this review.

      Substrate-targeting by the UPS and UIPS. Over 90% of the human proteome is regulated by the UPS[15]. These substrates include a vast array of structured (or folded) proteins, intrinsically disordered proteins (IDPs) and proteins containing intrinsically disordered regions (IDRs). Structured proteins must be unfolded prior to their degradation and can therefore only be cleared by the UPS[16,17]. Essentially, folded proteins cannot fit through the narrow axial pore of the 20S, making them inaccessible to degradation by the UIPS particles[18]. However, IDPs and IDR-containing proteins, which lack this three-dimensional structure, are thought to readily traverse the α-ring gate[19]. Twenty percent of cellular proteins are classified as IDPs and as many as 41% of the eukaryotic proteome is predicted to contain IDRs[20,21], suggesting that the substrate pool of the UIPS may be considerably large. These substrates are particularly relevant for this discussion because they include the proteins that accumulate in neurodegenerative disorders, such as amyloid beta, tau, TDP-43 and α-synuclein (**Table 1.1**)[22,23].

      In cells, IDPs typically have shorter half-lives relative to structured proteins[24]. The UPS and UIPS have both been shown to facilitate the rapid proteasomal degradation of

IDPs, such as p53 and p73[25]. Recognition of the IDPs by these pathways is mediated, in part, by disordered regions that act as signals (or degrons)[24]. Evidence for the UIPS in this process comes from experiments in which removal of the ubiquitinated lysine has been found to have little effect on turnover[26,27]. Thus, it seems that both the UPS and UIPS can contribute to the turnover of IDPs.

Role of the Proteasome in Ageing and Neurodegenerative Diseases

Neurodegenerative disorders, such as Alzheimer's disease (AD), Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS), are characterized by the progressive structural and functional impairment of neurons, resulting in neuronal death[28]. Although these diseases have different clinical symptoms, they are all associated with the accumulation of aggregated proteins[29,30]. This observation implicates the age-dependent decline in proteostasis and the contributing role of proteasomal dysfunction[31]. Consistent with this idea, species with increased longevity and long-lived individuals within a given species exhibit higher proteasome activity and are less susceptible to diseases, including neurodegeneration[32,33].

Many of the proteins implicated in neurodegeneration are IDPs[34]. Due to their conformational flexibility, IDPs are prone to aggregation[35] and present a particular risk in aged neurons, where the protein quality control network has deteriorated[36]. Against this backdrop, the proteasome likely plays a critical role because it rids neurons of damaged and misfolded proteins that may cause neuronal dysfunction. The role of the UPS in this process has received substantial attention due, in part, to the early observation that ubiquitinated proteins are found within disease-associated aggregates in postmortem

6

brains[31]. Indeed, many recent reviews have focused on the connection between the UPS and neurodegeneration[37]. Similarly, the immunoproteasome, composed of alternative subunits[38], likely plays a role, especially in the immune regulation of the diseases. The structure-function of the immunoproteasome has also been reviewed[39] and will not be discussed here. Instead, we focus on the relatively under-studied position of the UIPS.

Cells employ the UIPS to cope with proteotoxicity. During neurodegeneration, neurons are challenged with a higher substrate load, potentially causing an imbalance in proteostasis and creating a feed-forward loop that compromises proteasome function[41]. Specifically, increased levels of misfolded proteins are thought to "clog" the proteasome and impair its function[42]. There are several proposed mechanisms for this inhibition, including stabilization of the proteasome in its closed-gate conformation[43,44]. The feed-forward aspect of the relationship then comes about when the resulting proteotoxicity triggers mitochondrial dysfunction and increases reactive oxygen species (ROS) and DNA damage[45]. These processes contribute to further protein misfolding and inactivation of the proteasome through direct oxidation of its subunits[46]. Cells have adopted approaches to counteract this problem; for example, a rise in ROS induces expression of PA28$\alpha\beta$, which promotes UIPS activity[47]. Additionally, oxidative stress induces expression of 20S subunits through the oxidative stress sensor, Nrf2 (SKN-1)[48]. Interestingly, the levels of the UPS-associated PA700 remain constant under those conditions[49], suggesting that oxidatively damaged proteins may be preferentially cleared by the UIPS. Consistent with this idea, oxidative stress also mediates the Ecm29-dependent disassembly of 26S (**Figure 1.2A**)[50]. Collectively, these findings suggest that challenged cells rely on the UIPS to degrade oxidatively damaged and unfolded proteins.

7

Aged cells contain a latent pool of free (*e.g.* unbound) 20S proteasome. What are the relative contributions of the free 20S, the UPS and the UIPS to proteotoxicity and neurodegeneration? The answers are not yet clear, but some clues come from reports that the relative levels of 20S and PA-bound 20S change with ageing and disease. Using label-free proteomics in nine different human cell lines, Fabre *et al* estimates that 21 – 35% of the total proteasome pool is PA700-bound (26S) and less than 10% is bound to UIPS-specific PAs, while the remaining ~66% is unbound 20S[51] (**Figure 1.2B**). This is an interesting result because it suggests that cells may contain a latent pool of 20S proteasome that is not bound to any PA. Because PAs enhance the rate of proteolysis by as much as 20-fold[52], these findings suggest that some cells have a pool of 20S that is poised to be activated. How does this ratio change during ageing and disease? It is known that this downregulation is mediated, in part, by diminished expression of proteasome genes themselves[54]. To date, the distribution of PA-bound and unbound 20S has not been surveyed in a neurodegeneration model, but the ratio of unbound (*e.g.* free 20S) proteasome to UPS-specific 26S has been reported to increase with age. For example, in cultured human fibroblasts from old individuals, the amount of 20S proteasome is decreased by 2.5-fold whereas the level of PA700 is lowered by 6.5-fold, compared to fibroblasts from young individuals[55]. Conversely, the levels of PA28 remained unchanged,[55] suggesting a relative switch from UPS to UIPS pathways during ageing (**Figure 1.2B**). Although it is difficult to conclude a direct cause-and-effect, this increased availability of 20S roughly coincides with a decrease in total proteasome activity in many cells, tissues, and organisms during ageing[53].

Proteasomal upregulation accelerates the clearance of pathogenic proteins.

Together, these findings suggest that decreased proteasome activity, driven by a combination of reduced expression of 20S and an increase in free (*i.e.* non-PA bound) proteasome, may contribute to neurodegeneration. Accordingly, this model suggests that boosting proteasome activity may counteract the process. This goal could theoretically be achieved in a number of ways. For example, delivery of purified 20S to cells through direct injection has been shown to accelerate clearance of tau[56]. Similarly, cells expressing a 20S mutant, in which the gating N-termini of the α-subunits are deleted, are also partially protected from tau aggregation[57]. Pharmacologically, Finley *et al* identified small molecules that inactivate the deubiquitinating activity of USP14, the PA700-associated DUB, allosterically activate proteasomal degradation of polyUb-conjugated proteins and, consistent with the model, accelerate turnover of tau *in vitro*[58,59]. These findings have motivated others to search for molecules that directly bind the 20S to promote its activity (see below). Such a strategy might take advantage of the fact that only <10% of total proteasomes in a neurodegenerative disease model are intact 26S (**Figure 1.2B**). Consistent with this idea, over-expression of PA28 enhances protein clearance in ageing models[60]. The pharmacological equivalent of this approach would be to activate the 20S, in the absence of a PA. In the next sections, we discuss the opportunities and challenges of this possibility.

Pharmacological Activators of the 20S Proteasome

The proteasome has long been the subject of studies to identify inhibitors. Early work identified natural products, such as lactacystin, that inactivate it by mimicking

peptide substrates and covalently modifying the peptidase-containing β-subunits[61]. These compounds became widely used chemical probes and, subsequently, a few were approved for the treatment of multiple myeloma patients[53], where they seem to exploit the need of cancer cells for high protein turnover. More recently, other strategies have been developed to create inhibitors with different mechanisms[62]. Activators of the proteasome, in contrast, have received less attention; as mentioned above, such molecules would ideally bind to the unbound 20S and open its pore, acting as "artificial activators" to compensate for the loss of natural PAs and proteasome activity during ageing (**Figure 1.3**). However, the size and complexity of the 20S coupled with an incomplete understanding of the mechanisms of activation, has made the discovery of activators a difficult task. Furthermore, the field of 20S activators lacks the serendipitous discovery of natural product leads (such as lactacystin) as starting points. The next sections introduce the early chemical efforts to create proteasome activators and point out the substantial challenges that remain in the discovery, optimization and deployment of these molecules. For clarity, the sections are divided based on the chemical composition of the compounds.

Denaturants. The first reported proteasome activator was the detergent, sodium dodecyl sulfate (SDS), which was identified *in vitro* by measuring the chymotryptic activity of the 20S against the fluorogenic peptide substrate, succinyl-Leu-Leu-Val-Tyr-7-amino-4-methylcoumarin (LLVY-amc). Addition of SDS increases the turnover of LLVY-amc by 20-fold, but at levels above its critical micellar concentration. This result suggests that SDS acts as a detergent, likely activating by partially denaturing the ⍺-ring gate and allowing access of substrate into the pore. Another sign of this mechanism is that

stimulation occurs within a narrow concentration range: lower concentrations activate turnover, while higher levels inhibit activity, likely by more extensive denaturation of the 20S[63,64]. Similar trends are seen for polycations (*e.g.* polylysine), polyanionic lipids (*e.g.* cardolipin & heparin), fatty acids (*e.g.* oleic, linoleic, & linolenic acids) and the natural product oleuropein[65–69] (**Table 1.2A and B**). Molecules with this type of concentration-activity signature are likely to be rather non-specific activators, with mechanisms that make them difficult to optimize.

   Small molecules. Betulinic acid is a triterpene natural product that stimulates the 20S at low micromolar concentrations[70]. Unlike the detergents, betulinic acid preferentially stimulates the chymotryptic activity of the 20S (and not the other two activities)*.* However, attempts to improve the potency of betulinic acid through medicinal chemistry have only yielded analogs that inhibit (rather than stimulate) proteasome activity, suggesting complex structure-activity relationships (SAR)[71]. More recent reports have turned to high-throughput chemical screens to identify alternative molecules. These assays often rely on measurement of proteasome activity through the hydrolysis of LLVY-amc, while secondary assays are used to measure degradation of biologically relevant substrates[72,73]. For example, Trader *et al* screened a subset of the NIH Clinical Collection (NCC) library to discover two compounds, AM-404 and MK-886, that stimulate the 20S *in vitro* and enhance turnover of α-synuclein by 3 to 4-fold in cells[72]. In a parallel screen of the Natural Product Library (NPL) of the NCC, Coleman & Trader identified three additional molecules (denoted NPL-1, 2, & 3) that activate the 20S at low micromolar concentrations[74] (**Table 1.2C**). Beyond the specific molecules identified, one of the highlights of that work is the introduction of a rigorous suite of assays to triage non-specific

11

mechanisms. However, additional medicinal chemistry and structural biology efforts will be needed to advance chemical probes with the desired selectivity and potency.

In a parallel effort, Jones *et al* screened the NCC and Prestwick libraries, revealing chlorpromazine as a putative agonist of the 20S. This compound stimulated the activity of the 20S *by* 20-fold *in vitro*. Unbiased docking of chlorpromazine to the 20S predicted possible interactions with the intersubunit pocket of the α-ring, reminiscent of the natural HbYX motif. Thus, this chemical series may take advantage of intrinsic allosteric mechanisms to promote turnover. Chlorpromazine is a well-known dopamine D2 agonist, but chemical modifications suggested that the SAR for proteasome activation was distinct. However, the potency of these compounds remains limited (**Table 1.2C**)[73].

Peptides. Another strategy for stimulating the proteasome is to directly create mimics of the HbYX motif. Indeed, peptides inspired by the HbYX-containing C-termini of PAs can stimulate the 20S *in vitro* at mid-micromolar concentrations[5]. Moreover, the activity of these HbYX peptides seems to depend on which subunits of PA700 they originated from[75,76]. Some of these HbYX peptides can enhance proteasomal degradation of a model substrate and rescue the 20S from inhibition by toxic amyloid-β oligomers[43]. While promising, peptide-based activators of the 20S have intrinsic challenges that will need to be overcome. For example, they have poor membrane permeability, uncertain selectivity and typically low metabolic stability. Despite these hurdles, the recently reported crystal structure of the HbYX peptide-bound eukaryotic 20S[77] might provide a structure-guided way to advance this approach. It is important to note that any compounds that bind this site would expectedly displace the UPS and UIPS-specific PAs.

Thus, even though they might stimulate the pool of free 20S, their pharmacology is expected to be complex.

Conclusion and Prospectus

There is mounting evidence that decreased activity of the proteasome contributes to neurodegeneration. Although decreased 26S is assuredly a key component of this decline, we have focused here on the contributions of the UIPS and the free 20S. This focus is based on evidence that both the PA28-bound and free 20S seem to become more prominent during ageing. Moreover, many of the neurodegenerative disease-associated proteins, such as tau, are IDPs and thus particularly good substrates for these complexes.

The increased level of free 20S during ageing has interesting implications for drug discovery. Specifically, molecules that boost the function of this "low activity" 20S pool might partially compensate for diminished proteasomal function. While a compelling idea, there are major challenges to pursuing this concept and the current molecules are not yet up to this task. One challenge is that many compounds can have detergent-like activity that is potent, but ultimately untenable for the creation of activators. Thus, any screening effort is highly likely to produce denaturation-like artifacts that must be removed through subsequent secondary assays and careful analysis of their dose dependence. In addition, one cannot use the typical approach of adding Tween or Triton-X in screening buffers to minimize the discovery of aggregators or pan-assay interference (PAINS) molecules[78]. Thus, additional types of artifacts are likely to populate the list of apparent "hits". These

two challenges are expected to result in higher-than-normal failure rates during compound triage and special care will be needed to select high quality scaffolds.

References

1. Smith DM, Benaroudj N, Goldberg A. Proteasomes and their associated ATPases: A destructive combination. *J Struct Biol*. **2006**; 156(1):72-83.

2. Goldberg AL. Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy. *Biochem Soc Trans*. **2007**; 35(1):12 LP-17.

3. Chen LH and CH. Proteasome Regulators: Activators and Inhibitors. *Curr Med Chem*. **2009**; 16(8):931-939.

4. Kish-Trier E, Hill CP. Structural Biology of the Proteasome. *Annu Rev Biophys*. **2013**; 42(1):29-49.

5. Smith DM, Chang SC, Park S, Finley D, Cheng Y, Goldberg AL. Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's α Ring Opens the Gate for Substrate Entry. *Mol Cell*. **2007**; 27(5):731-744.

6. Finley D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annu Rev Biochem*. **2009**; 78(1):477-513.

7. Finley D, Ulrich HD, Sommer T, Kaiser P. The ubiquitin-proteasome system of Saccharomyces cerevisiae. *Genetics*. **2012**; 192(2):319-360.

8. DeMartino GN, Moomaw CR, Zagnitko OP, et al. PA700, an ATP-dependent activator of the 20 S proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. *J Biol Chem*. **1994**; 269(33):20878-20884.

9. Glickman MH, Rubin DM, Fried V A,Finley D. The regulatory particle of the Saccharomyces cerevisiae proteasome. *Mol Cell Biol*. **1998**; 18(6):3149-3162.

10. Schmidt M, Hanna J, Elsasser S, Finley D. Proteasome-associated proteins: Regulation of a proteolytic machine. *Biol Chem*. **2005**; 386(8):725-737.

11. Sadre-Bazzaz K, Whitby FG, Robinson H, Formosa T, Hill CP. Structure of a Blm10 Complex Reveals Common Mechanisms for Proteasome Binding and Gate Opening. *Mol Cell*. **2010**; 37(5):728-735.

12. Li J, Rechsteiner M. Molecular dissection of the 11S REG (PA28) proteasome activators. *Biochimie*. **2001**; 83(3-4):373-383.

13. Förster A, Masters EI, Whitby FG, Robinson H, Hill CP. The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol Cell*. **2005**;18(5):589-599.

14. Strickland E, Hakala K, Thomas PJ, DeMartino GN. Recognition of misfolded proteins by PA700, the regulatory subcomplex of the 26S proteasome. *J Biol Chem*. **2000**;275(8):5565-5572.

15. Kwon YT, Ciechanover A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem Sci*. **2017**;42(11):873-886.

16. Rock KL, Gramm C, Rothstein L, et al. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*. **1994**;78(5):761-771.

17. Craiu A,Gaczynska M, Akopian T, et al. Lactacystin and clasto-lactacystin beta-lactone modify multiple proteasome beta-subunits and inhibit intracellular protein degradation and major histocompatibility complex class I antigen presentation. *J Biol Chem*. **1997**;272(20):13437-13445.

18. Hagai T, Levy Y. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc Natl Acad Sci*. **2010**;107(5):2001-2006.

19. Orlowski M, Wilk S. Ubiquitin-independent proteolytic functions of the proteasome.

Arch Biochem Biophys. **2003**;415(1):1-5.

20. Baugh JM, Viktorova EG, Pilipenko E V. Proteasomes can degrade a significant proportion of cellular proteins independent of ubiquitination. *J Mol Biol*. **2009**;386(3):814-827.

21. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome informatics*. **2000**;11:161-171.

22. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*. **1996**;35(43):13709-13715.

23. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*. **2008**;37(1):215-246.

24. van der Lee R, Lang B, Kruse K, et al. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep*. **2014**;8(6):1832-1844.

25. Asher G, Tsvetkov P, Kahana C. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73 service A mechanism of proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev*. **2005**;19:316-321.

26. Peña MMO, Xing YY, Koli S, Berger FG. Role of N-terminal residues in the ubiquitin-independent degradation of human thymidylate synthase. *Biochem J*. **2006**;394(1):355-363.

27. Wiggins CM, Tsvetkov P, Johnson M, et al. BIMEL, an intrinsically disordered protein, is degraded by 20S proteasomes in the absence of poly-ubiquitylation. *J Cell Sci*. **2011**;124(6):969-977.

28. Popovic D, Vucic D, Dikic I. Ubiquitination in disease pathogenesis and treatment. *Nat Med*. **2014**;20(11):1242-1253.

29. Walker LC, LeVine H. The cerebral proteopathies: neurodegenerative disorders of protein conformation and assembly. *Mol Neurobiol*. **2000**;21(1-2):83-95.

30. Sherman MY, Goldberg AL. Cellular defenses against unfolded proteins: A cell biologist thinks about neurodegenerative diseases. *Neuron*. **2001**;29(1):15-32.

31. Dantuma NP, Bott LC. The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution. *Front Mol Neurosci*. **2014**;(7):1-18.

32. Ruby JG, Smith M, Buffenstein R. Naked mole-rat mortality rates defy gompertzian laws by not increasing with age. *Elife*. **2018**;7:1-18.

33. Chondrogianni N, Petropoulos I, Franceschi C, Friguet B, Gonos ES. Fibroblast cultures from healthy centenarians have an active proteasome. *Exp Gerontol*. **2000**;35(6-7):721-728.

34. Deger JM, Gerson JE, Kayed R. The interrelationship of proteasome impairment and oligomeric intermediates in neurodegeneration. *Aging Cell*. **2015**;14(5):715-724.

35. Levine ZA, Larini L, LaPointe NE, Feinstein SC, Shea J-E. Regulation and aggregation of intrinsically disordered peptides. *Proc Natl Acad Sci*. **2015**;112(9):2758-2763.

36. Ciechanover A, Brundin P. The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron*. **2003**;40(2):427-446The ubiquitin proteasome system in neurodeg.

37. Dennissen FJA, Kholod N, van Leeuwen FW. The ubiquitin proteasome system in

neurodegenerative diseases: Culprit, accomplice or victim? *Prog Neurobiol*.
2012;96(2):190-207.

38. Kniepert A, Groettrup M. The unique functions of tissue-specific proteasomes.
Trends Biochem Sci. **2014**;39(1):17-24.

39. Ferrington DA, Gregerson DS. *Immunoproteasomes: Structure, Function, and
Antigen Presentation*. **2012**.

40. Uversky VN. Alpha-synuclein misfolding and neurodegenerative diseases. *Curr
Protein Pept Sci*. **2008**;9(5):507-540.

41. Huang Y, Mucke L. Alzheimer mechanisms and therapeutic strategies. *Cell*.
2012;148(6):1204-1222.

42. San Martín Á, Rodriguez-Aliaga P, Molina JA, Martin A, Bustamante C, Baez M.
Knots can impair protein degradation by ATP-dependent proteases. *Proc Natl Acad
Sci*. **2017**;114(37):201705916.

43. Thibaudeau TA, Anderson RT, Smith DM. A common mechanism of proteasome
impairment by neurodegenerative disease-associated oligomers. *Nat Commun*.
2018.

44. Bennett EJ, Bence NF, Jayakumar R, Kopito RR. Global impairment of the ubiquitin-
proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion
body formation. *Mol Cell*. **2005**;17(3):351-365.

45. Lin MT, Beal MF. Mitochondrial dysfunction and oxidative stress in
neurodegenerative diseases. *Nature*. **2006**;443(7113):787-795.

46. Reichmann D, Voth W, Jakob U. Maintaining a Healthy Proteome during Oxidative
Stress. *Mol Cell*. **2018**;69(2):203-213.

47. Li J, Powell SR, Wang X. Enhancement of proteasome function by PA28 overexpression protects against oxidative stress. *FASEB J*. **2011**;25(3):883-893.

48. Li X, Matilainen O, Jin C, Glover-Cutter KM, Holmberg CI, Blackwell TK. Specific SKN-1/NrF stress responses to perturbations in translation elongation and proteasome activity. *PLoS Genet*. **2011**;7(6):9-11.

49. Pickering AM, Staab TA, Tower J, Sieburth D, Davies KJA. A conserved role for the 20S proteasome and Nrf2 transcription factor in oxidative stress adaptation in mammals, Caenorhabditis elegans and Drosophila melanogaster. *J Exp Biol*. **2013**;216(4):543-553.

50. Wang X, Chemmama IE, Yu C, et al. The proteasome-interacting Ecm29 protein disassembles the 26S proteasome in response to oxidative stress. *J Biol Chem*. **2017**;292(39):16310-16320.

51. Fabre B, Lambour T, Garrigues L, et al. Label-free quantitative proteomics reveals the dynamics of proteasome complexes composition and stoichiometry in a wide range of human cell lines. *J Proteome Res*. **2014**;13(6):3027-3037.

52. Masson P, Lundin D, Söderbom F, Young P. Characterization of a REG/PA28 proteasome activator homolog in dictyostelium discoideum indicates that the ubiquitin- And ATP-independent REGγ proteasome Is an ancient nuclear protease. *Eukaryot Cell*. **2009**;8(6):844-851.

53. Schmidt M, Finley D. Regulation of proteasome activity in health and disease. *Biochim Biophys Acta*. **2014**;1843(1):13-25.

54. Lee C, Klopp RG, Weindruch R, Prolla TA. Gene Expression Pro le of Aging and Its

Retardation by Caloric Restriction. *Science (80-)*. **1999**;(285):1390-1393.

55. Hwang JS, Hwang JS, Chang I, Kim S. Age-associated decrease in proteasome content and activities in human dermal fibroblasts: restoration of normal level of proteasome subunits reduces aging markers in fibroblasts from elderly persons. *J Gerontol*. **2007**;62(5):490-499.

56. Han DH, Na H-K, Choi WH, et al. Direct cellular delivery of human proteasomes to delay tau aggregation. *Nat Commun*. **2014**;5:5633.

57. Choi WH, de Poot SAH, Lee JH, et al. Open-gate mutants of the mammalian proteasome show enhanced ubiquitin-conjugate degradation. *Nat Commun*. **2016**;7:10963.

58. Lee B-H, Lee MJ, Park S, et al. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature*. **2010**;467(7312):179-184.

59. Boselli M, Lee BH, Robert J, et al. An inhibitor of the proteasomal deubiquitinating enzyme USP14 induces tau elimination in cultured neurons. *J Biol Chem*. **2017**;292(47):19209-19225.

60. Vilchez D, Saez I, Dillin A. Organismal Ageing and Age-Related Diseases. *Nat Commun*. **2014**;5:1-13.

61. Kim KB, Myung J, Sin N, Crews CM. Proteasome inhibition by the natural products epoxomicin and dihydroeponemycin: Insights into specificity and potency. *Bioorg Med Chem Lett*. **1999**;9(23):3335-3340.

62. McDaniel TJ, Lansdell TA, Dissanayake AA, et al. Substituted quinolines as noncovalent proteasome inhibitors. *Bioorganic Med Chem*. **2016**;24(11):2441-2450.

63. Orlowski N, Wilk S. A multicatalytical protease complex from pituitary that forms

enkephalin and enkephalin containing peptides. *Biochem Biophys Res Commun*. **1981**;101(3):814-822.

64. Wilk S, Orlowski M. Evidence that Pituitary Cation-Sensitive Neutral Endopeptidase Is a Multicatalytic Protease Complex. *J Neurochem*. **1983**;40(3):842-849.

65. Ohkubo I, Gasa S, Namikawa C, Makita A, Sasaki M. Human Erythrocyte multicatalytic proteinase: Activation and binding to sulfated galacto- and lactosylceramides+. *Biochem Biophys Res Commun*. **1991**;174(14):1133-1140.

66. Ruiz de Mena I, Mahillo E, Arribas J, Castaño JG. Kinetic mechanism of activation by cardiolipin (diphosphatidylglycerol) of the rat liver multicatalytic proteinase. *Biochem J*. **1993**;296:93-97.

67. Watanabe N, Yamada S. Activation of 20S proteasomes from spinach leaves by fatty acids. *Plant Cell Physiol*. **1996**;37(2):147-151.

68. Tanakas K, Yoshimura T, Ichihara A, et al. A High Molecular Weight Protease in the Cytosol of Rat Liver. *J Biol Chem*. **1986**;261(32):15204-15207.

69. Katsiki M, Chondrogianni N, Chinou I, Rivett AJ, Gonos ES. The olive constituent oleuropein exhibits proteasome stimulatory properties *in vitro* and confers life span extension of human embryonic fibroblasts. *Rejuvenation Res*. **2007**;10(2):157-172.

70. Kashiwada Y, Hashimoto F, Cosentino LM, Chen CH, Garrett PE, Lee - KH. Betulinic acid and dihydrobetulinic acid derivatives as potent anti-HIV agents. *J Med Chem*. **1996**;39(5):1016-1017.

71. Huang L, Ho P, Chen CH. Activation and inhibition of the proteasome by betulinic acid and its derivatives. *FEBS Lett*. **2007**;581(25):4955-4959.

72. Trader DJ, Simanski S, Dickson P, Kodadek T. Establishment of a suite of assays

that support the discovery of proteasome stimulators. *Biochim Biophys Acta*. **2017**;1861(4):892—899.

73. Jones CL, Njomen E, Sjögren B, Dexheimer TS, Tepe JJ. Small Molecule Enhancement of 20S Proteasome Activity Targets Intrinsically Disordered Proteins. *ACS Chem Biol*. **2017**; 12(9):2240-2247.

74. Coleman RA, Trader DJ. Development and Application of a Sensitive Peptide Reporter to Discover 20S Proteasome Stimulators. *ACS Comb Sci*. **2018**; acscombsci.7b00193.

75. Gillette TG, Kumar B, Thompson D, Slaughter C a., DeMartino GN. Differential roles of the COOH termini of AAA subunits of PA700 (19 S regulator) in asymmetric assembly and Activation of the 26 S proteasome. *J Biol Chem*. **2008**; 283(46):31813-31822.

76. Kim Y-C, DeMartino GN. C Termini of Proteasomal ATPases Play Nonequivalent Roles in Cellular Assembly of Mammalian 26 S Proteasome. *J Biol Chem*. **2011**; 286(30):26652-26666.

77. Witkowska J, Gizyńska M, Grudnik P, et al. Crystal structure of a low molecular weight activator Blm-pep with yeast 20S proteasome - Insights into the enzyme activation mechanism. *Sci Rep*. **2017**; 7(1):1-11.

78. Baell JB, Nissink JWM. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017 - Utility and Limitations. *ACS Chem Biol*. **2018**; 13(1):36-44.

A. Components of the proteasome degradation pathway.
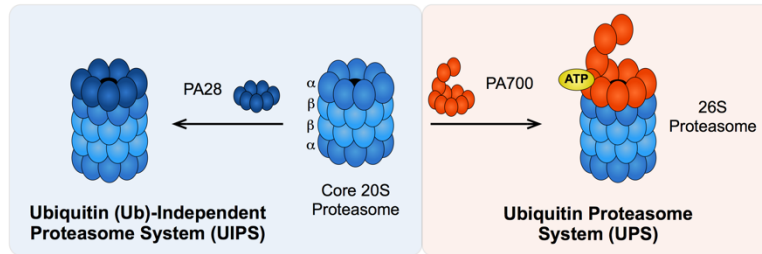
B. Diagram of the degradative fates of native and modified cellular proteins.
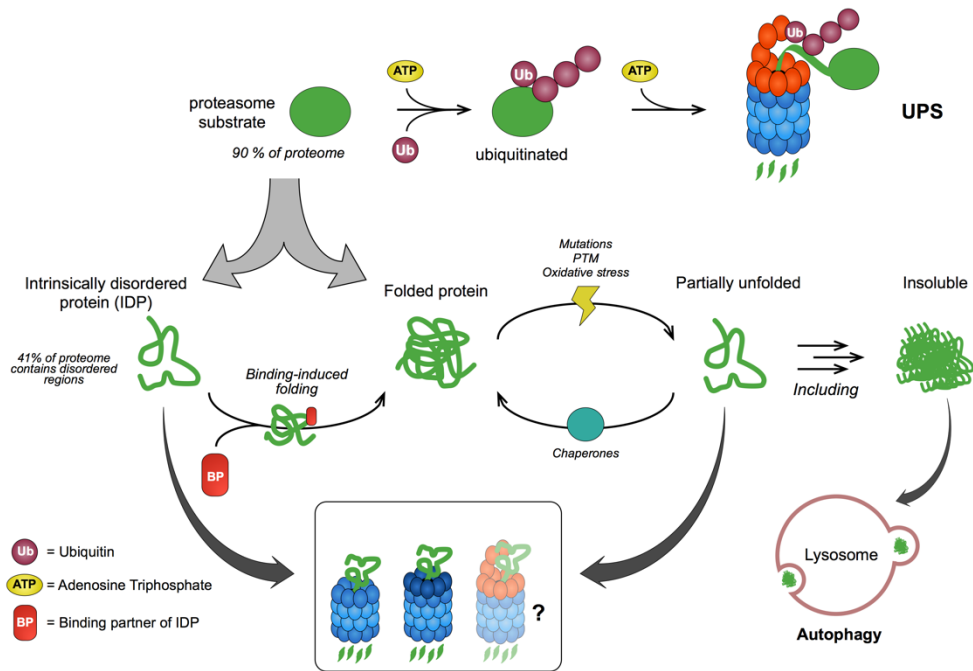
Figure 1.1. Structure, function, and substrate-profile of the proteasome degradation pathway.
A, Components of the ubiquitin proteasome system (UPS) and the ubiquitin (Ub)-independent proteasome system (UIPS). **B**, Diagram of substrate turnover by the different proteasome pathways. The 26S proteasome degrades over 90% of the proteome, including intrinsically disordered proteins (IDPs) and folded proteins, by the Ub- and ATP-dependent UPS. The UIPS targets substrates independent of Ub-conjugation and can effectively degrade IDPs, which constitute up to 41% of the proteome, but not folded proteins due to their three-dimensional structure. Cellular stresses, including mutations, post-translational modifications (PTMs) and oxidative damage, can partially unfold structured proteins making them susceptibility to turnover by the UIPS. Partially unfolded proteins (and IDPs) can self-associate to form insoluble aggregates, which cannot be processed by the proteasome, and are predominantly cleared from the cell through the lysosome-autophagy pathway.

A. Venn diagram of age-associated/neurodegenerative effects on the proteasome.

Stress on UIPS
- Oxidative stress-induced upregulation of PA28*
- Oxidatively damaged protein

- Increased substrate load (disease-associated & misfolded proteins)
- Oxidative modification
- Protein aggregates
- Autophagy dysfunction
- Genetic downregulation of 20S

Stress on UPS
- Oxidative stress-induced disassembly of 26S
- Genetic downregulation of PA700/19S
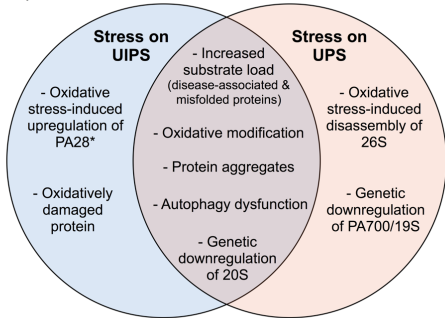
*Changes enhance proteasome function

B. Diagram of the redistribution of proteasome complexes in ageing.

Age-associated and neurodegenerative stress

Young
21 - 35% UPS
~ two-thirds of 20S remain unbound

Cellular Proteasome Composition

Old
10% UPS
90% UIPS

Upregulation of proteasome activity
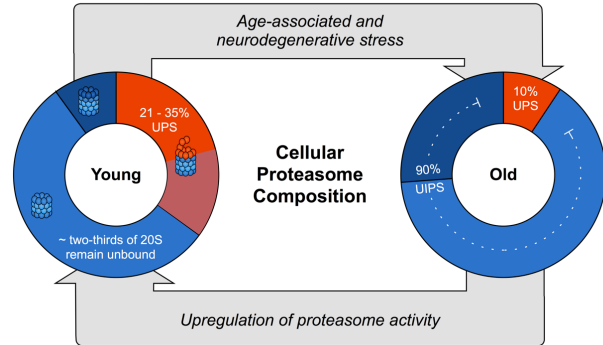
Figure 1.2. Proteasomal regulation in ageing and neurodegenerative disease models.
A, Partial list of stresses that affect UPS (orange) and UPS (blues) function. **B**, Summary of the cellular composition of proteasome complexes for UPS (26S) and UIPS (20S and PA28-20S complex) in young versus aged cells based on studies in fibroblasts[51, 55]. Note that this simplification does not account for the contribution of the 26S in degrading substrates in a ubiquitin-independent manner.
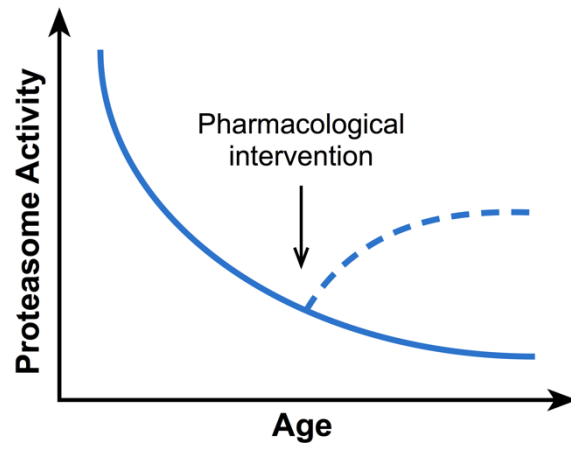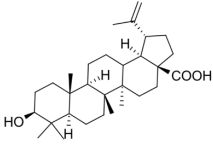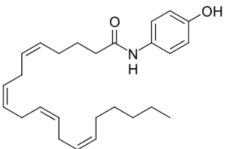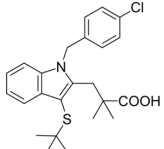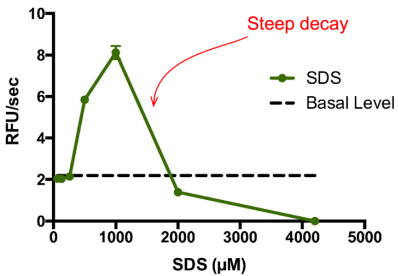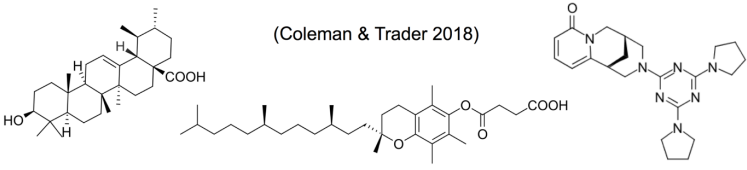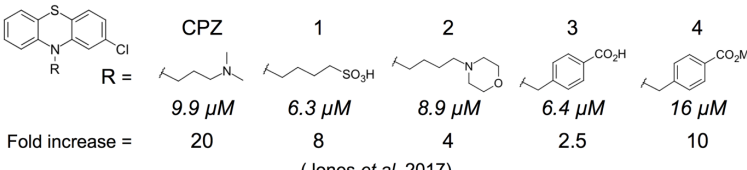
Figure 1.3. Schematic representation of the general decline in proteasome activity during ageing. In theory, small molecules that partially restore proteasome function can potentially alleviate ageing.

Table 1.1. IDPs and associated neurodegenerative diseases. Adapted from Uversky[40].

Protein	Disorder by prediction (%)[a]	Function	Disease(s)
Amyloid-β	16.9	Peptidic fragment of APP, which regulates synapse formation, and neuronal plasticity	Alzheimer's disease Amyloidosis
Tau	77.6	Promote the assembly of and stabilizes neuronal microtubules	Tauopathies Alzheimer's disease Corticobasal degeneration Pick's disease Progressive supranuclear palsy
TDP-43	57.3	Transcriptional repression, pre-mRNA splicing, and translational regulation	Amyotrophic lateral sclerosis and frontotemporal lobar degeneration
α-Synuclein	90.7	Regulate synaptic vesicles	Alzheimer's disease (α) Multiple system atrophy (α) Parkinson's disease (α, β, γ) Diffuse Lewy body disease (α, β, γ)
β-Synuclein	87.3		
γ-Synculein	100		
FUS	90.7	Transcriptional regulation (initiation & repression), and RNA-binding	Amyotrophic lateral sclerosis

[a] Disorder was predicted by PONDR VSL2.

Table 1.2. Activators of the 20S proteasome.

A. Non-Specific Activators	C. Small Molecule Activators (EC_{50})

A. Non-Specific Activators

Sodium dodecyl sulfate (SDS)
Polycations (*e.g* polylysine)
Polyanions (*e.g* cardolipin & heparin)
Fatty acids (*e.g* oleic & linolenic acid)
Oleuropein

B. Sample dose-response curve for denaturants

(Trader *et al.* 2017)

(A) Denaturant-like activators. (B) SDS. (C) Measured EC50 values of 20S proteasome chymotryptic activity determined *in vitro*. The maximum fold increase over vehicle control was also calculated for chlorpromazine (CPZ) analogues[69-74].

C. Small Molecule Activators (EC_{50})

Betulinic acid	AM-404	MK-866
5.5 µM	*32 µM*	*32 µM*
(Huang *et al.* 2007)	(Trader *et al.* 2017)	

(Coleman & Trader 2018)

NPL-1	NPL-2	NPL-3
14 µM	*7 µM*	*10 µM*

	CPZ	1	2	3	4
R =	*9.9 µM*	*6.3 µM*	*8.9 µM*	*6.4 µM*	*16 µM*
Fold increase =	20	8	4	2.5	10

(Jones *et al.* 2017)

28

Chapter 2

The YΦ motif defines the structure-activity relationships of human 20S proteasome activators

Kwadwo A. Opoku-Nsiah, Andres H. de la Pena, Sarah K. Williams, Nikita Chopra,

Andrej Sali, Gabriel C. Lander and Jason E. Gestwicki

Abstract

The proteasome degradation system facilitates turnover of most eukaryotic proteins within its core 20S complex. To enable substrate entry into the 20S proteasome (20S), gating loops must be displaced through binding of HbYX motifs that are present at the C-termini of proteasome activators (PAs). The HbYX motif has been predominantly characterized in the archaeal 20S, whereas little is known about the sequence preferences of the human 20S (*h*20S). Here, we synthesized and screened ~120 HbYX-like peptides, revealing unexpected differences from the archaeal system and defining the *h*20S recognition sequence as the Y-F/Y (YΦ) motif. To gain further insight, we created a functional chimera of the optimized sequence, NLSYYT, fused to the model activator, PA26[E102A]. A cryo-EM structure of PA26[E102A]-*h*20S identified key interactions, including non-canonical contacts and gate-opening mechanisms. Finally, we demonstrated that the YΦ sequence preferences are tuned by valency, allowing multivalent PAs to sample greater sequence space. These results expand the model for termini-mediated gating and provide a template for the design of pharmacological activators of the *h*20S.

Introduction

The proteasome is a critical regulator of protein homeostasis that degrades ~90% of all eukaryotic proteins[1]. The enzymatic activity of this system is carried out by the 20S proteasome (20S), a cylindrical, complex composed of four stacked rings enclosing an axial channel. To be degraded, potential substrates must first diffuse through a pore at the center of the distal α-rings before encountering the peptidase sites within the inner β-

rings. N-terminal extensions of the α-subunits "gate" entry into the pore, limiting the degradation of bystander proteins[2]. In turn, this barrier creates a key regulatory role for the proteasome activators (PAs)[3–5], large particles that bind the 20S, open the gates, and facilitate substrate selection and entry. One evolutionarily conserved[6–8] way that PAs achieve this goal is by using a tripeptide motif at their extreme C-termini, which is characterized by a hydrophobic amino acid, followed by a tyrosine and then any amino acid (HbYX). HbYX motifs open the gates by docking into pockets located between adjacent α-subunits of the 20S (termed α-pockets) (**Figure 2.1A**)[9–11]. Unlocking the details of HbYX recognition at this protein-protein interaction (PPI) will deepen our understanding of the gate opening mechanism and enable novel strategies to regulate protein degradation in cells.

The HbYX model for gate opening was pioneered from studies by Goldberg and colleagues using the C-terminal sequence of the archaeal PA, proteasome-activating nucleotidase (PAN)[9]. Such work revealed key roles for the penultimate Tyr residue and the terminal carboxylate of the HbYX motif, which make hydrogen-bonding contacts within the α-pockets of the archaea 20S[12,13]. However, incongruities in the eukaryotic proteasome system, such as its heteromeric assemblies, suggest that evolutionary changes may lead to distinct interactions governing gating from archaea to the human 20S (*h*20S). Profiling the sequence preferences unique to humans has been difficult within native proteasome complexes because the interacting PPIs are complex and yet constrained. In particular, the topology of PA700/19S, the eukaryotic homolog to PAN, limits each of its six distinct C-termini to a specific α-pocket of the 20S[14]. The contributions of these cognate PPIs, are convoluted by binding cooperativity and confounding effects

from ATP hydrolysis[15–17] and allosteric effects from additional subunits of PA700[18–20]. One way to circumvent these issues, and directly probe the fundamental structure-function relationships, has been to use synthetic peptides instead of native PAs. Indeed, HbYX-containing peptides derived from the C-termini of the human Rpt subunits of PA700, have been shown to bind the α-pockets and stimulate turnover of substrates by the 20S[9,21,22]. Inspired by these findings, we hypothesized that peptides could be used to understand the sequence preferences of the HbYX motif in the *h*20S. Importantly, further investigation into this key PPI will inform the design of small molecules that bind the α-pockets as a means of regulating 20S activity[23].

In this study, we designed and synthesized ~120 peptides derived from Rpt5's C-terminus and evaluated their ability to stimulate the peptidase activity of the *h*20S *in vitro*. This analysis revealed sequence preferences that considerably differed from the canonical HbYX motif as derived in archaea. We refer to this re-defined preference as the nYΦn motif. To better understand the structural underpinnings of the YΦ motif, we grafted an optimal sequence, NLSYYT, to the C-termini of an inert PA platform and solved a 2.9 Å resolution structure of the PA bound to the *h*20S by cryo-electron microscopy (cryo-EM). Remarkably, the same orientation of NLSYYT was observed in five of the seven α-pockets, suggesting a conserved pharmacophore. Analysis of the bound YΦ motif revealed specific inter- and intra-molecular contacts, which were biochemically validated with peptides to be involved in molecular recognition and gate opening. Finally, using a series of chimeras, we found that the valence of the PA displaying the YΦ motif (*e.g.* monomer *vs.* heptamer) tuned the sequence preferences, with multivalent PAs able to overcome otherwise non-ideal sequences. Together, these studies characterize termini-

dependent gate opening in the human proteasome and establish a model that is inclusive of monovalent activators of the 20S.

Results

Peptides derived from *h*Rpt5 dose-dependently activate the *h*20S *in vitro*. Throughout this work, we will use nomenclature in which the carboxy-terminal residue of the HbYX motif is termed P1, the next residue P2, *etc*. In this parlance, the canonical HbYX model is defined as having a hydrophobic residue at P3, a Tyr at P2, and any residue at P1. Furthermore, it is known that the terminal carboxylate of the P1 residue forms a critical salt bridge with a conserved cationic side chain at the base of the α-pocket, for example αLys66 of the *Thermoplasma acidophilum* 20S (**Figure 2.1A**)[9]. We will default to numbering the using the residue numbering of the *T. acidophilum* 20S. To determine if molecular recognition in the *h*20S also conforms to the HbYX model, we synthesized, characterized, and assayed a library of peptides derived from the C-terminus of the human Rpt5 (*h*Rpt5), with the native sequence: ANLQYYA. For each peptide, its capacity to dose-dependently accelerate proteasomal turnover was monitored by the hydrolysis of a standard fluorogenic substrate (suc-LLVY-amc) *in vitro*. Because sodium dodecyl sulfate (SDS) and other detergent-like molecules have been reported to non-specifically activate the 20S, these otherwise useful additives are typically excluded from buffers in proteasome activity assays[24]. Peptide-based activators have also been shown to exhibit atypical dose-responses, including partial to full inhibition of the 20S[25], likely through non-specific mechanisms.

To minimize the discovery of false-positives in our *h*Rpt5 peptide screen, we initially sought to optimize buffer conditions and establish a pipeline to detecting peptides that non-specifically activate the *h*20S. We referenced previous reports to determine a buffer composition that would increase the dynamic range of 20S activity between basal and stimulated proteasome samples[5,26,27]. Using this buffer, we observed that high concentrations of either peptide or established small-molecule activators would inhibit substrate turnover of the *h*20S. Supplementing the assay buffer with a non-ionic surfactant (0.01% Pluronic F-68 ®) seemed to restore dose-responsive for most activators, suggesting that solubility may influence peptide activity (**Figure S2.1A**, **B, C, & D**). We noted that peptides containing reactive residues, such as Cys and Met, lost the capacity to stimulate the *h*20S over time (**Figure S2.1E**), and were therefore excluded from this study. We also omitted peptides that were visibly insoluble at 10 mM in DMSO (**Figure S2.1F**). DMSO-solubilized peptides, particularly those with hydrophobic sequences, have the propensity to form colloidal aggregates when resuspended in an aqueous buffer[28], which can potentially influence proteasome function in the activity assay. Using dynamic light scattering (DLS) to detect aggregation, we flagged peptides suspected to form colloidal aggregated and their activities were closely evaluated before including any data in the analysis pipeline.

In the first series of *h*Rpt5 peptides, we performed an alanine mutational scan of ANLQYYA. Consistent with previous reports[21], the parent peptide (termed wildtype or WT) stimulated turnover of suc-LLVY-amc by ~3× (**Figure 2.1B**). An EC_{50} value could not be determined, due to relatively weak potency and limited solubility, so, at this stage, we compared hydrolysis rates. Replacing either the P3 ('Hb') or P2 ('Y') positions with Ala

completely ablated activity, in agreement with the HbYX model[9]. Surprisingly, replacing the P4, P5 or P6 residues with Ala also had appreciable effects, including an unexpected enhancement in activity by Ala substitution at P4 (1.3× over WT), complete ablation of activity at P5, and modest impairment at P6 (~80% of WT) (**Figure 2.1B**). A series of N-terminal truncations revealed that the 6- and 7-mer peptides were significantly more stimulatory than shorter ones (**Figure 2.1C**), further suggesting a functional role for residues "upstream" of the tripeptide motif. These results advised us to interrogate the contributions of the HbYX residues in addition to residues N-terminal of the tripeptide motif.

Structure-activity relationships (SAR) of *h*Rpt5 peptides deviate from the HbYX model. Next, we designed a series of hexapeptides to probe the structure-activity relationships (SAR) in more detail. Peptides were synthesized and N-terminally acetylated (Ac), which generally enhanced activity over the free amine (**Figure S2.2A**). Acetylated peptides were then assessed at a single dose (250 µM) against the *h*20S to obtain an initial overview of relative activity. In the first set of comparisons, we varied the P1, P2, and P3 positions, replacing the wild-type residue with chemically and/or structurally diverse amino acids. Preliminarily, we reported that the *h*20S had some sequence preference at the P1 position, with Thr favored over other residues that were tested. More importantly, Tyr seemed to be exclusively required at the P3 position; such that even other hydrophobic residues, including Leu and Ile, were not tolerated. Lastly, either Phe or Tyr was preferred at the P2 position (**Figure 2.1D**). Thus, these results demonstrate that differences exist between the SAR in the *h*20S and the archaeal 20S (**Figure S2.3**)[9].

We were particularly intrigued by the gain in activity caused by Ala at P4 (see **Figure 2.1B**), so we generated a focused series of P4-substituted hexapeptides to explore this position further. In this collection, the P1 was uniformly replaced with Thr, which we found to improve aqueous solubility and activate better than Ala, allowing us to obtain saturable stimulation curves and calculate EC_{50} values (**Figure S2.4A**). We found that none of the P4 substitutions completely ablated activity, suggesting that, unlike P2 or P3, the side chain of the P4 residue indirectly regulates gate opening. Nonetheless, modifications at P4 impart the greatest modulatory effect on hydrolysis (up to 3× over WT), with positively charged or bulky side chains generally exhibiting lesser stimulation (**Figure 2.1E**). Additionally, residues known to disrupt α-helical character (*e.g.* Gly and Pro) were well-tolerated at the P4 position[29]. Lastly, we explored the contributions of the P5 and P6 position. Even though we had previously found that truncations of these residues diminished activity, substitutions of P5 or P6 had modest effects (see **Figure S2.2B, C, & D**), suggesting that the side chain identity was less important in these positions.

Combining the highlights of the SAR, we incorporated optimal P1 and P4 substitutions to probe whether they were synergistic. In these experiments, we were particularly interested in understanding which substitutions might enhance the apparent EC_{50} (as a pseudo approximation of affinity) and which ones impact the rate of hydrolysis. The results showed that Thr at P1 improved the EC_{50}, while Ser at P4 enhanced both EC_{50} and rate of turnover. Combining both substitutions had an additive effect (**Figure 2.1F**), yielding an optimized sequence Ac-NLSYYT (Opt5; EC_{50} = 35.5 ± 1.1 μM). Contrary to the HbYX model, the P2 position accepted either Phe or Tyr (F/Y – denoted

as Φ) to suggest that H-bonding of the P3 Tyr was required over that of a P2 Tyr. To test this idea, we substituted either the P2 or P3 Tyr residues of Opt5 (referred to here as Opt5^{YY} for clarity) with Phe (Opt5^{YF} or Opt5^{FY}, respectively) and measured their ability to stimulate h20S. Expectedly, removal of the P3 hydroxyl in Opt5^{FY} dramatically reduced activity by ~70%. Loss of the P2 hydroxyl had no appreciable effect on turnover but improved the potency of Opt5^{YF} (EC$_{50}$ = 17.3 ± 1.1 μM) by 2-fold relative to Opt5^{YY} (**Figure 2.1F**). To confirm that Phe generally improved the potency of peptides, we assessed a P2 Φ residue in two additional peptide sequences (Ac-NLSYΦA and Ac-NLGYΦT) and noted up to a 3-fold improvement in potency (**Figure S2.4B**). Previous work has established precedence for a functional HbYX motif containing a P2 Phe[30], suggesting that the Φ side chain of the P2 residue may undergo additional contacts to promote gate opening.

Taken together, our biochemical data reports that proteasome activation by hRpt5 peptides occurs through C-terminal interactions of at least the last four residues. The sequence requirements differ from the HbYX motif and are summarized herein as the nYΦn (or YΦ) model. Modifications from hRpt5 to Opt5 stimulated suc-LLVY-amc turnover along with the tryptic-like activity of the h20S, which was assayed with boc-LRR-amc (**Figure S2.5A**). These improvements in activity, including the P2 Phe, were also noted in assays measuring the hydrolysis of the nonapeptide, open-gate sensitive substrate (FAM-LFP)[31] (**Figure 2.5B**). These findings imply that Opt5 stimulates the h20S by opening the α-ring gate.

PA26$^{E102A-Opt5}$ induces terminus-dependent gate opening of the h20S. To understand the molecular basis for the sequence preferences in hRpt5 as detailed in the

YΦ motif we attempted to determine the structure of *h*20S bound to Opt5 by cryo-EM. Unfortunately, we observed that Opt5 bound with inconsistent occupancy across the α-pockets and were thus discouraged from pursing a structure of the Opt5-*h*20S complex. Rather, inspired by previous studies[8,13,30,32], we sought to use the homo-heptameric activator PA26 as a scaffold for Opt5. Instead of C-terminal HbYX motifs, PA26 has an activation loop that directly repositions the reverse-turn loop at αPro17 in the α-ring, displacing the N-terminal gating residues (**Figure 2.2A**)[3,4]. An Ala substitution in the activation loop (E102A) renders PA26 inactive[8]. Previous studies showed that the capacity of the disabled-loop mutant (PA26[E102A]) to induce gate opening was rescued by grafting PAN's HbYX motif in place of the native C-termini[13,30]. Thus, we envisioned using PA26[E102A] as a scaffold to enable high-resolution structural studies of terminus-dependent gate opening in the *h*20S.

PA26 from archaea is capable of activating the human 20S[33], however, it remains unclear whether the E102A mutation impairs this function , as has been reported in archaea. Using the *Trypanosoma brucei* PA26, we verified that PA26[E102A] was unable to stimulate turnover by the *h*20S. Moreover, PA26[E102A] could no longer bind *h*20S, as measured by biolayer interferometry (BLI) (**Figure 2.2B**). Next, we installed Opt5 in place of the last six native residues of PA26 (PA26[E102A-Opt5]) to generate a PA that now bound the *h*20S (K_D = 360 ± 160 pM) and potently stimulated its activity (EC_{50} = 25.5 ± 1.1 nM) (**Figure 2.2B**). To probe how Opt5 binds and activates *h*20S, we then determined the cryo-EM structure of the singly capped *h*20S-PA26[E102A-Opt5] complex (2.9 Å in overall resolution) (**Figures. S2.6, S2.7 and Table S2.2**). A single state was resolved with the liganded α-ring displaying N-terminal extensions that were displaced from the central pore

(Figure 2.2C). The diameter of the pore was widened by 3.8 Å relative to the unliganded α-ring of the h20S (**Table S2.3**), which aligned with other published open-gate structures[8]. PA26$^{E102A-Opt5}$ did not induce allosteric gate opening of the distal α-ring, consistent with a previous study on the analogous mutant PA26 and archaeal 20S complex[32]. Taken together, these findings suggest that step-wise and allosteric gate opening in the 26S proteasome is likely attributed to conformational dynamics of PA700 and is not intrinsic to the HbYX mechanism[17].

Cryo-EM resolves open-gate conformation of the h20S. Using this structure, we first characterized the gate opening mechanism. The open-gate conformation is typically regulated by a cluster of conserved N-terminal residues (αTyr8, αAsp9, αPro17, and αTyr26)[7,8]. Specifically, these residues are re-positioned during gate opening and then anchored by a characteristic set of intra- and intermolecular contacts, which stabilize the N-termini away from the central pore. However, these residues are less conserved in humans, particularly in α1 and α2 (α-subunits are labeled to match the numbering used in yeast 20S), but the implications to gating have yet to be described. In the liganded α-ring of our structure, the N-termini of α5, α6, and α7 formed ordered clusters at the α-subunit interface, consistent with the fully open state[20,34]. This includes the positioning of αPro17 within CH-π distance from αTyr26 and αTyr8. However, the clusters typically formed at the α1/α2 interface were poorly resolved, suggesting a disordered state potentially caused by missing contacts between the non-canonical α1 Phe8 and α2 Ser9 residues (**Figure S2.8A**). The destabilizing effect also seemed to propagate from the α1/α2 cluster adjacent α-subunits resulting in the N-terminal extensions of the α2, α3, and α4 partially obstructing the pore, which has previously been observed by 26S

proteasome structures[16,17]. These observations suggest that termini-mediated *h*20S activation occur within the context of asymmetric, open-gate α-rings (**Figure S2.8B**).

To assess the conformational changes associated with gate opening, we aligned the β-ring of our structure with the apo *h*20S (PDB ID: 4R3O)[35], because this region is known to be relatively unchanged during activation[4,13]. We observed that the Cα atoms of all seven αPro17 residues were radially displaced between 1.0 and 3.7 Å during gate opening. Consistent with the analyses of the N-terminal gate positions, the "partially open" α1- and α2-subunits had the smallest shifts in Pro17 (1.0 and 1.3 Å, respectively) (**Figure S2.8C**). The modest movement of α2 in the *h*20S contrasts to the relatively large shifts typically observed by the α2's N-terminus in yeast[30]. This is interesting because the yeast α2, like the human α2, also has a non-canonical Ser9. Thus, it is possible that α1 Phe8 in the *h*20S, not α2 Ser9, is more restrictive of the fully open 20S conformation.

We next examined how these gate opening mechanisms might be mediated by the termini. First, we confirmed that the disabled activation loop of PA26[E102A-Opt5] does not interact with αPro17, suggesting that, as intended, this structure would reveal only termini-specific mechanisms. At the α-pockets, most of the P2 and P3 Tyr side chains of bound Opt5 made H-bond contacts with residues that are located in the reverse turn between αPro17 and helix 0 (**Figure S2.9A**). This finding suggested that residues within Opt5 worked directly through αPro17 to mediate gate opening. As further support of this idea, the α-subunits in these Opt5-bound pockets were rotated an average of 1.5 ± 0.4° about the axial channel (**Figure S2.9B**). Previous studies have attributed such rigid body rotation in the α-ring to binding of a C-terminal HbYX motif[12,13,17,32]. In addition, we

observed side chain rearrangements throughout the α-pockets that suggest induced-fit conformational changes upon binding of Opt5 (**Figure 2.3A**), which has also been implicated in termini-induced gate opening[13]. Overall, activation of the *h*20S proceeds through the conserved gating mechanism; however, these analyses illustrate subtle distinctions of the open-gate α-ring between the yeast and human 20S that are potentially linked to the non-canonical α1 Phe8 of the human α-ring.

Opt5-bound α-pockets reveal key contacts for terminus-dependent *h*20S activation. To further elucidate mechanisms of the YΦ motif, we next examined the interactions of Opt5 at the C-terminal regions (C-tails) of PA26[E102A-Opt5] in the *h*20S α-pockets. This analysis was aided by the matched stoichiometry of C-tails to α-pockets, together with the fact that the α-pockets were locally resolved to better than 2.9 Å resolution (see **Figure S2.7D**). Six of the seven α-pockets contained well-defined density corresponding to the C-tails of PA26[E102A-Opt5]. As previously predicted[8,20,36], the α7/α1-pocket (interfacing the α7 and α1 subunits) was unable to bind Opt5, likely because it does not have a canonical Lys residue (α1 His68). Of the remaining six sites, one of them, the C-tail in the α1/α2-pocket, was noticeably distinct. Specifically, its terminal carboxylate was displaced (5.1 Å) from the ε-amino group of α1 Lys63 and therefore unable to form the salt bridge, which has previously been shown to be critical for binding of native PAs and HbYX peptides[9,12]. To verify the importance of the negatively charged terminal carboxylate, we amidated the C-terminus of Opt5 peptide and confirmed that it could no longer stimulate the *h*20S (**Figure 2.1F**). Hence, we categorized the C-tails in the α2/α3, α3/α4, α4/α5, α5/α6 and α6/α7 pockets as binding in an "anchored" (A) state

and the C-tail in the $\alpha1/\alpha2$-pocket as binding in the "unanchored" (U) state (**Figure 2.3A, B**).

The peptide backbone of Opt5, of both A and U states, were well-resolved for the P1, P2, P3 and P4 positions, enabling further assessment of the $Y\Phi$ motif and the implications of these two binding states. The backbone conformations of the C-tails are also different between the A and U states. In particular, the dihedral angles of the P2-P3 residues of the Opt5 C-tail form a β-strand structure for the A state but a right-handed α-helix for the U state (**Figure 2.3C**). In a recent study of the archaeal 20S-P complex, the terminal residues of PAN in PA26[E102A-PAN] adopt α-helices when docked into α-pockets of the *T. acidophilum* 20S[32], suggesting that the β-strand conformation might be specific to eukaryotic 20S proteasomes and/or $Y\Phi$ motifs. The five C-tails that bound in the A state are nearly identical; each predicted to make between five and nine hydrogen bonds in the α-pocket, including the critical salt bridge (**Figure 2.3A**). There are fewer contacts (3 H-bonds and no salt bridge) made between the U state C-tail and the $\alpha1/\alpha2$-pocket. Another distinction between the U and A states is the configuration of predicted H-bonds between the $Y\Phi$ motif and residues in the α-pocket, most notably, αGlu25. In the U state, H-bonding occurs between the carboxy side chain of $\alpha1$ Glu26 and the hydroxyl of the P2 Tyr. In the A state, hydrogen bonding is mediated between αGlu25 and the P3 Tyr side chain, which allows for additional contacts to the α-pocket by the hydroxyl group of the P2 Tyr (*i.e. $\alpha2/\alpha3$, $\alpha4/\alpha5$, & $\alpha5/\alpha6$*) (**Figure 2.3B**). Conservation of αGlu25 from *T. acidophilum* to humans, including across all seven distinct α-subunits (see **Figure S2.8A**), further highlights the utility of targeting this H-bond contact. Thus, these structural

and functional studies suggest that polar contacts within the α-pocket, namely H-bonding between the P3 Tyr and αGlu25, are involved in termini-mediated gate opening of the *h*20S.

Intramolecular stacking of P3 and P2 residues regulates α-pocket engagement. Considering the heterogeneity of the α-pockets in the hetero-oligomeric *h*20S, we were struck by how the C-tails adopted strikingly similar orientations. Indeed, when we overlay all six bound C-tails from the PA26[E102A-Opt5]-*h*20S structure, the P2 and P3 Tyr residues adopt a nearly identical, stacked orientation (see **Figure 2.3C**). Moreover, we noted that prior structures of eukaryotic PA-20S complexes also report YΦ motif-containing C-tails in a similar stacked arrangement[10,37] (**Figure S2.10**). In our structure, the average distance (*R*) between the centroids of the P2 and P3 Tyr rings (4.5 ± 0.2 Å) and the average angle (*θ*) between the ring planes (26.8 ± 3.3°) closely match theoretical predictions for off-center parallel, π stacking (**Figure 2.4A**)[38]. Thus, we postulated that intramolecular π stacking facilitates termini-dependent gate opening by pre-orienting Opt5 to a competent conformation within the α-pocket. To test this hypothesis, we designed non-acetylated peptide analogs of Opt5[YF] that perturbed π stacking and measured their ability to stimulate the *h*20S. A full dose-response curve revealed an 80% reduction in activity relative to Opt5[YF] by installing a cyclohexyl group (4) at the P2 position (Opt5[Y4]), which can no longer π stack with the P3 Tyr side chain (**Figure 2.4C**). Likewise, an inversion of the aromatic quadrupole through a pentafluoro phenylalanine (5) at the P2 position (Opt5[Y5]) significantly reduced potency (EC$_{50}$ > 600 µM) and activity (by 80%) relative to Opt5[YF]. Less drastic perturbation of the π electron density, through installation of ortho- or meta-chloro groups (2 and 3; **Figure 2.4B**), had

43

intermediate consequences (**Figure 2.4C**). Although we cannot rule out steric effects, these experiments suggest that intramolecular π-stacking interactions between the Y and Φ residues contribute to Opt5 activity, perhaps by pre-orienting the motif for engagement with αGlu25.

Valency tunes the sequence preferences for proteasomal activation. Native PAs and other 20S-binding partners are often multivalent, displaying multiple C-tails from a single scaffold. One effect of multivalence is that binding avidity is increased, which seems to promote cellular assembly of proteasome complexes[18,26,27,39,40]. However, a less appreciated effect of valency in biological recognition is that it also tunes specificity[41], amplifying some preferences and allowing more variability in others. To determine if valency influences the sequence preferences of terminal recognition by the h20S, we generated and tested additional chimeras of the heptavalent PA26[E102A]. In these studies, we focused on the P2 and P3 positions because of their aforementioned, key roles. Unlike the monovalent peptides, Phe substitution of the P3 Tyr did not affect the rate of hydrolysis nor potency of PA26[E102A-Opt5] (PA26[YY] & PA26[FY]; EC_{50} = 25.5 ± 1.1 & 22.6 ± 1.1 nM, respectively). Interestingly, Phe substitution of the P2 Tyr slightly improved the EC_{50} (PA26[YF]; 15.1 ± 1.1 nM) (**Figure 2.5A**), as was shown with the Opt5 peptides. We also noted that PA26[YY] has a k_{on} that is ~3x slower than that of PA26[YF] and PA26[FY]; and yet, the calculated K_D for PA26[YY] falls within error of the others, owing to its significantly slower k_{off} (**Figure S2.11B**). We then generated an alternate, monovalent PA by fusing peptide sequences onto the C-terminus of maltose-binding protein (MBP). Using these PAs, we found that MBP with a C-terminal Opt5 (MBP[YY]) stimulated the h20S (EC_{50} = 41.8 ± 1.2 µM) equipotently to Opt5[YY] peptide. Likewise, the P2 Phe mutant was more

potent (MBP^{YF}; EC$_{50}$ = 10.0 ± 1.2 µM), mirroring the SAR from the monovalent peptides. The P3 Phe mutant exhibited diminished activity (MBP^{FY}; EC$_{50}$ > 400 µM), however, its relative activity was significantly lower than its peptide equivalent to suggest that Opt5^{FY} is capable of adopting a gating-competent binding pose to stimulate the h20S **(Figure 2.5B)**. Thus, valency appeared to allow the h20S to be more permissive of missing contacts in the PA's C-tails; whereas monovalent ones adhere to the YΦ motif.

Next, we interrogated the impact valency had on the relative importance of intramolecular π stacking by generating P3 or P2 Ala mutations in all three PA-types: monovalent peptide, monovalent MBP, and multivalent PA26^{E102A}. In stimulation assays, Ala mutations at either P3 or P2 completely inactivated both peptides (Opt5^{AY} and Opt5^{YA}) and MBP chimeras (MBP^{AY} and MBP^{YA}) (**Figure S2.4D & G**). Without a P3 Tyr and intramolecular π stacking, monovalent PAs were unable to stimulate h20S which implies that the importance of the P3 Tyr seems to be contingent on π-stacking interactions that are attributed by aromaticity of the P2 side chain. In PA26^{E102A}, disrupting π stacking with a P3 Ala mutation had no effect on hydrolysis, however, the P2 Ala mutation ablated stimulation (PA26^{YA}; EC$_{50}$ > 3000 nM). Furthermore, PA26^{AY} stimulated the h20S more potently (EC$_{50}$ = 8.3 ± 1.2 nM) than the constructs capable of π stacking (**Figure S2.4D**), which suggests that multivalent PA's do not require either P3 Tyr or intramolecular π stacking to induce gate opening. Another distinction between mono- and multi-valent activators is the residue preferences at the P1 and P4 positions, which were effectively masked in the heptavalent PA26^{E102A}. For example, PA26^{NLQYYA} activates more potently (EC$_{50}$ = 6.2 ± 1.1 nM) than the improved upon Opt5 sequence (**Figure S2.4D**). These

results, along with published SAR of the hexavalent PAN[9], suggest the multivalent PAs activate according to the requirements previously established by the HbYX model.

Finally, we generated P2 and P3 Phe double mutants, which precludes both H-bonds while preserving aromaticity to enable π stacking, in the three PA-types and assessed activity. In PA26^{E102A}, the Phe mutant (PA26^{FF}) was a partial agonist of the human 20S (relative activity = 50%) and stimulated at a lower potency (EC$_{50}$ = 56.2 ± 1.2 nM) than every active C-terminally modified construct (**Figure S2.4H**). Relative to the full agonists, PA26^{FF} exhibits a weaker binding affinity (decreased by 100×); however, the K_D values are equivalent between the partial agonist (PA26^{FF}; 35 ± 18 nM) and the inactive PA26^{YA} (36 ± 9 nM) (**Figure S2.11A**). This implies that the binding kinetics of PA26^{YA} should enable partial gate opening according to the activity exhibited by PA26^{FF}. These findings suggest that the terminus of PA26^{YA} binds in a gating-incompetent state whereas the C-terminal sequence of PA26^{FF}, notably its intramolecular π stacking, favorably orients the motif in the α-pocket and opens the gate. However, the Phe double mutation was not enough for Opt5 nor MBP to activate the h20S (Opt5^{FF} and MBP^{FF}; EC$_{50}$ > 400 μM) (**Figure S2.4B & C**), demonstrating that monovalent PAs require both π stacking and P3 tyrosyl H-bond interactions. Collectively, these data show that valency reduces the stringency for termini-mediated activation of the human 20S (**Figure 2.5C**).

Discussion

An updated HbYX model for binding the α-pockets and initiating h20S gate opening. Pioneering studies of the HbYX motif in the archaeal proteasome (**Figure 2.6A**)[9,12,13], has contributed significantly to our understanding of proteasome activation

and 20S regulation. Herein, we aimed to deepen that knowledge through characterization of how Rpt5 peptides stimulate the human 20S. This effort uncovered narrower preferences for the P1 'X' and P3 'hydrophobic' residues and an unanticipated role for the P4 position, suggesting that the h20S exhibits greater sequence selectivity than the archaeal system. Considering a subset of amino acids, only Ala and Thr were strongly preferred at P1, in contrast to broad specificity exhibited in the archaeal system (see **Figure S2.3**). A reason for this preference in the h20S was revealed by our structural studies, which showed that Thr in this position engages in H-bonding within some α-pockets. The structural rationale for the preference at P4 is less clear. Tentatively, small, aliphatic residues might be preferred over bulkier residues due to potential steric clashes with a neighboring β-strand-turn in the α-pockets. Likewise, a subset of polar amino acids, such as Ser, Thr, and Asp were also relatively preferred (see **Figure 2.1E**), indicating that H-bonding in that region may be involved. However, further work is needed to elucidate the importance of these interactions in gate opening. Finally, another key observation was that a Tyr was prioritized at P3, whereas the P2 position accepted Φ. Our structural and proteasome activity analyses revealed novel intermolecular H-bonds and intramolecular π-stacking interactions of the P2 and P3 side chains as putative mediators of gate opening. Together, these structure-function observations led us to propose the YΦ motif as an alternative to the HbYX model for the h20S (**Figure 2.6B**).

In our h20S-PA26$^{E102A\text{-}Opt5}$ structure, we noted that the C-tail binds in two different states (A and U), and thus postulated that Opt5 peptides likely occupy the α-pockets in either of these distinct conformations. We also noted differences in the intermolecular polar contacts mediated between Opt5 and the α-pocket, which suggest that binding in

the A state preferentially stabilizes gate opening over that of the U state. Herein, we speculate how the A and U states potentially represents putative docking poses for Opt5$^{Y\Phi}$ (Opt5^{YF} and Opt5^{YY}) and Opt5^{FY} peptides, respectively. Specifically, the diminished activity of Opt5^{FY} can be linked to fewer H-bond contacts and the displaced salt bridge depicted in the U state. Similarly, Opt5$^{Y\Phi}$ peptides bind in the A state and form more contacts in the α-pocket, including the salt bridge and H-bonds between the P3 Tyr and αGlu25 side chains, to stabilize the open-gate conformation. In reported structures of both yeast[20] and human[37] 26S proteasomes, the C-tail of the Rpt3 subunit binds the α1/α2-pocket with a displaced salt bridge while adopting an α-helical turn as observed in the U state. Moreover, the P3-P2 residues of Rpt3's C-terminal consensus sequence are, like Opt5^{FY}, F-Y, which we predicted to be associated with the U state. It is plausible that the observed U state is entirely independent of Opt5 and its interactions and instead stems from unique features of the α1/α2-pocket. We require additional structures, particularly a peptide-bound structure of the activated proteasome complex, to deepen our understanding of the YΦ gating mechanism.

Valency tunes selectivity for the C-terminal PA sequences. Using a series of monovalent (free peptide and MBP chimera) and heptameric (PA26^{E102A}) scaffolds, we observed how valency tunes the sequence preferences of the C-tails. More specifically, the multivalent PA was able to overcome a subset of the requirements suggested by the YΦ motif, most strikingly, a Tyr residue at the P3 position. This finding implies that the canonical HbYX model may not sufficiently describe the sequence preferences for all PAs, particularly the monovalent, eukaryotic activator PA200/Blm10[10]. As a first step in asking that question, we aligned the C-terminal sequences of naturally occurring PAs

across multiple organisms (**Table S2.4**). Indeed, PA200 ended in residues that closely match the YΦ motif (**Figure 2.7A**). Moreover, a recent structure revealed that the C-terminus of PA200 adopts both the β-strand and the intramolecular off-centered π stacking[42], and affirmed that the specifications of the YΦ motif, including intermolecular contacts, is potentially associated with a gating-competent state. Thus, we postulate that monovalent PAs will strictly adhere to the YΦ model as characterized in the peptide study.

We then aligned the C-terminal sequences of multivalent PAs, focusing on the archaeal PAN and the three Rpt subunits containing HbYX motifs (Rpt2, Rpt3, & Rpt5) of PA700 from multiple organisms. This exercise showed that the C-terminus of Rpt5 closely resembled the YΦ motif, except that the P3 position accommodates a Phe residue, and the P4 position samples a wider range of residues (**Figure 2.7B**). Notably, Rpt5 is the most stimulatory of the Rpt subunits[21,22], underscoring the important influence of the YΦ motif on *h*20S activation, even in the context of some multivalent PAs. The less stimulatory Rpt2 and Rpt3 subunits, in contrast, displayed more divergent sequences from the YΦ motif, most notably because Tyr was not conserved at the P3 position (**Figure 2.7C & D**). Even less consensus was observed when we aligned the C-termini of available PAN sequences. Specifically, the P2 tended to be Φ while the residues at the remaining positions seemed arbitrary (**Figure 2.7E**). Unlike PA700, PAN is homo-oligomeric and so no individual monomer (*e.g.* Rpt5) is predicted to make an out-sized contribution to overall binding or stimulation. Thus, in the case of PAN, valency seems to largely override the requirements of both the YΦ and the more permissive HbYX motifs. Non-adherence to the canonical HbYX model was similarly noted in an examination of another hexavalent regulator of the archaeal 20S, Cdc48[40]. Together, these observations

suggest that C-tail sequence identity, plus a major contribution from valency, combine to dictate the biomolecular interactions with the proteasome. Additional binding partners of the 20S, with HbYX motifs of varying sequences and valences, have been identified in eukaryotes but the full extent to which they regulate the proteasome remains limited[40,43–45]. Hence, it may be useful to create sub-categories of HbYX-like models, in which the valency of the PA is a key, defining characteristic.

Implications for the discovery of pharmacological proteasome activators. Loss of proteasome function is implicated in many devasting proteinopathies, including cardiomyopathy and neurodegenerative disorders[46]. Consistent with this idea, boosting 20S activity, by introducing either native[47] or engineered activators[48], has been found to be protective in cell-based disease models. These observations have motivated campaigns to discover drug-like molecules that mimic the activity of PAs[23]. The current work defines the SAR associated with h20S α-pockets, provides a template for the rational design of pharmacological proteasome activators, including those previously reported by the Tepe and Trader Groups[49,50]. For instance, the binding kinetic studies with the PA26 chimeras suggest that longer dwell times within the α-pockets promote stabilization of the open-gate conformation. We suspect that pre-arrangement of the YΦ motif, through π stacking, minimizes the entropic costs of binding for Opt5. For molecules that conform to this pharmacophore model, we should be able to drive potency by rigidification of the bioisosteric equivalent to the Tyr-Tyr π stack. We also found that Opt5 docked into 6 of the 7 available α-pockets with remarkable conformational identity. Although Rpt peptides have been confirmed to target these α-pockets[36], a high-resolution structure of peptide-bound h20S is needed to fully ascertain the requisite occupancy and

binding interactions for inducing gate opening with a small molecule. And what about engagement of the final unoccupied $\alpha 7/\alpha 1$-pocket with the uncharged $\alpha 1$ His68? Occupancy of this α-pocket might require additional strategies, such as shifting the local pKa of the His to promote protonation of its side chain in order to facilitate salt bridge interactions. While such questions remain unanswered, the current work extends our knowledge of the mechanisms for activation of the *h*20S and provides a blueprint that details the requirements for designing novel proteasome activators.

Acknowledgements

Author contributions. K.A.O.-N. and J.E.G. designed the studies and wrote the manuscript. All authors edited the manuscript. K.A.O.-N., S.K.M., and N.C. conducted

biochemistry experiments, performed data analysis, and generated necessary reagents. A.H.P. performed cryo-EM sample preparation, data collection, and data processing. J.E.G., K.A.O.-N., A.S. and G.C.L provided funding.

Materials and Methods

Reagents. Human 20S proteasome was purchased from the Proteasome Center. pET28 6His PA26 (49V) cloning vector was a gift from Philip Coffino. pET 6His MBP TEV LIC cloning vector (1M) was a gift from Scott Gradia (Addgene plasmid # 29656).

Strains and Plasmids. The *E. coli* strain Top10 was used for propagating plasmids. BL21 (DE3) cells were used for expression and purification of recombinant proteins.

Peptide synthesis. Peptides were synthesized by Fmoc solid phase peptide synthesis on a Syro II peptide synthesizer (Biotage) at ambient temperature and atmosphere on a 12.5 µM using either pre-loaded Wang resin or Rink amide resin (Sigma-Aldrich). Coupling reactions were run with 4.9 eq. of HCTU (*O*-(1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluoro-phosphate), 5 eq. of Fmoc-AA-OH and 20 eq. of *N*-methylmorpholine (NMM) in 500 µl of *N,N*-dimethyl formamide (DMF). Fmoc-AA-OH was double coupled for 8 min while shaking for each position. Fmoc deprotection was conducted with 500 µl 40% 4-methylpiperidine in DMF for 3 min, followed by 500 µl 20% 4-methylpiperidine in DMF for 10 min and six washes with 500 µl of DMF for 3 min. Acetylation of the N-terminus was achieved by reacting 20 eq. acetic anhydride and 20 eq. *N,N*-diisopropylethylamine (DIPEA) in 1 mL DMF for 2 h while shaking. Peptides were cleaved with 500 µl of cleavage solution (95% trifluoroacetic acid (TFA), 2.5% water and

2.5% triisopropylsilane) while shaking for 2 h. Crude peptides were precipitated in 15 ml cold 1:1 diethyl ether: hexanes and air-dried overnight. Crude peptides were solubilized in a 1:1:1 mixture of DMSO: water: acetonitrile, filtered, and purified by high-performance liquid chromatography (HPLC) on an Agilent Pursuit 5 C18 column (5 mm bead size, 150 × 21.2 mm) using Agilent PrepStar 218 series preparative HPLC. The mobile phase consisted of A, 0.1% TFA in water and B, 0.1% TFA in acetonitrile. Peptides were purified to >95% homogeneity confirmed by liquid chromatography-mass spectrometry before solvent was removed by lyophilization. Peptides were resuspended in 1:1 water: acetonitrile, lyophilized again in tared tubes and stocks were stored at -20 °C.

Protein expression and purification in bacteria. All proteins were produced in *E. coli* BL21(DE3) and stored at −80 °C.

PA26 WT and PA26[E102A] (*Trypanosoma brucei*, 49V, His tagged) were expressed from a pET28 construct with a N-terminal 6His tag. *E. coli* were grown in terrific broth (TB) with the requisite antibiotic at 37 °C, induced with 1 mM IPTG in log phase, and grown for an additional 3 h. Cells were harvested by centrifugation, resuspended in His resin binding buffer (20 mM Tris pH 7.9, 20 mM NaCl, 10 mM imidazole) supplemented with protease inhibitors, sonicated and clarified by centrifugation. Clarified lysate was applied to Ni-NTA His-Bind Resin (Novagen). Resin was washed with binding buffer and then eluted in batches with buffer containing increasing amounts of imidazole (up to 300 mM). Purified protein was treated with 1 mM DTT for 1 h at ambient temperature and imidazole was removed by overnight dialysis into storage buffer (20 nM Tris pH 8.0, 200 mM NaCl) for later use.

Chimeric PA26^{E102A} (*Trypanosoma brucei*, 49V, His tagged) were expressed from a pMCSG7 construct with an N-terminal TEV-cleavable 6His tag. *E. coli* were grown in TB with the requisite antibiotic at 37 °C, induced with 1 mM IPTG in log phase, cooled to 18 °C and grown overnight. Cells were harvested by centrifugation, resuspended in His resin binding buffer (20 mM Tris pH 7.9, 20 mM NaCl, 10 mM imidazole) supplemented with protease inhibitors, sonicated and clarified by centrifugation. Clarified lysate was applied to Ni-NTA His-Bind Resin (Novagen). Resin was washed with binding buffer and then eluted in batches with buffer containing increasing amounts of imidazole (up to 800 mM). Purified protein was treated with 1 mM DTT for 1 hr at ambient temperature and imidazole was removed by overnight dialysis into storage buffer for later use.

MBP WT and chimeric MBP (His tagged) were expressed from pET and pMCSG7 constructs, respectively. *E. coli* were grown in TB supplemented with 0.2% glucose (instead of glycerol) with the requisite antibiotic at 37 °C, induced with 1 mM IPTG in log phase, cooled to 25 °C and grown overnight. Cells were harvested by centrifugation, resuspended in amylose resin binding buffer (20 mM Tris pH 7.8, 200 mM NaCl, 1 mM EDTA) supplemented with protease inhibitors, sonicated and clarified by centrifugation. Clarified lysate was applied to amylose resin (New England BioLabs, E8021S). Resin was washed with binding buffer and then eluted with 200 mM maltose in binding buffer. Purified protein was dialyzed into storage buffer for later use.

Proteasome activity assays. Stimulation of *h*20S proteasome (Proteasome Center) was assayed in 384-well plates (Greiner Bio-One, 781209) using the fluorogenic peptide substrates suc-LLVY-amc (AnaSpec, AS-63892), boc-LRR-amc (AdipoGen, AG-CP3-0014) or FAM-LFP (5-FAM-AKVYPYPMEK(QXL520)-NH2; AnaSpec) in assay

buffer containing 50 mM Tris pH 7.5,10 mM MgCl$_2$, 200 µM ATP, 1 mM DTT, and 0.01% Pluronic F-68® (Gibco Life Technologies, 24040032) in a total volume of 30 µL. Human 20S (final concentration of 4 nM) was incubated in the presence or absence of activators (peptides, 0.2 – 600 µM; MBP 0.2 – 400 µM; PA26, 0.04 – 3500 nM) at ambient temperature for 5 min. Substrate (suc-LLVY-amc, 10 µM; boc-LRR-amc, 20 µM; or FAM-LFP, 100 nM) was added immediately before reading. Fluorescence intensity of suc-LLVY-amc & boc-LRR-amc (excitation, 355 nm; emission, 440 nm; cutoff 435 nm) or FAM-LFP (excitation, 490 nm; emission, 520 nm; cutoff, 515 nm) were monitored at 30 - 60 s intervals for 30 min at 25 °C using a Spectramax M5 microplate reader (Molecular Devices). The hydrolysis rate was calculated from the slope of the curve between 100 and 500 s (500 and 1000 s for boc-LRR-amc) in arbitrary units (RFU per s).

Data were processed and fit in GraphPad Prism 8.0. The baseline hydrolysis was normalized to the total mean activity for the lowest concentration of every activator assayed in a given plate and the maximal hydrolysis was normalized as reported. Normalized activity were plotted relative to $\log_{10}$(activator). Data was fit to the model for log(agonist) versus response (variable slope). In equation (1), X = log[activator, µM] except for PA26-based activators, which are in nanomolar.

$$Y = \text{Bottom} + \frac{(\text{Top}-\text{Bottom})}{(1+10^{((\log\text{EC}_{50}-X)\,\times\,\text{Hillslope}))}}\ (1)$$

Dynamic Light Scattering. From 10 mM stocks in DMSO, three 4-fold serial dilutions of each peptide were prepared in assay buffer that was filtered with Millex-GS 0.22 µm sterile filter unit to final concentrations of 600, 150, and 37.5 µM. A final volume of 20 µL was added to a 384-well plate (Corning, 3540) and the unsealed plate was

55

centrifuged at 25 °C for 45 s at 500 x g. Colloidal aggregation was measured using a Wyatt Technologies DynaPro Plate Reader II (acquisition time of 2 s, 10 acquisitions, with auto attenuation, and temperature controlled at 25 °C).

Biotinylation of *h*20S. Commercially available *h*20S was dialyzed into PBS (Gibco Life Technologies) pH 7.5 and biotinylated with 20 eq of NHS-Biotin (Thermo Scientific, 20217) at 4 °C for 2 h. Excess reagent was removed with a PD-10 desalting column (Amersham Biosciences), separated into aliquots, snap-frozen, and stored at -80 °C for later use.

Binding kinetics analysis. Biolayer interferometry (BLI) data of PA26 activators were measured using an Octet RED384 (ForteBio). The reactions were carried out in black 384-well plates (Greiner Bio-One, 781209) at 25 °C with a volume of 85 μl per well in BLI buffer (assay buffer containing 0.2% (w/v) BSA (Sigma)). Biotinylated *h*20S (10 nM) were immobilized on streptavidin (SA) biosensor. Serial dilutions of PA26 in BLI buffer were used as analyte. Association was observed by immersing loaded biosensors into solutions of analyte for 450 s. No binding was detected of the analyte to unloaded sensors. Dissociation was observed by transferring the sensor to a well containing binding buffer and no analyte for 1200 s. Affinity (K_D) and kinetic parameters (k_{on} and k_{off}) were calculated from a global fit (2:1 heterogenous ligand model) of the data using the Octet data analysis software.

Cryo–electron microscopy grid preparation. h20S proteasome and PA26[E102A-Opt5] activators were mixed by dilution into a buffer system (20 mM Tris pH 7.5, 20 mM NaCl, 10 mM KCl, 1 mM DTT, and 0.025% Nonidet P40) to a final concentration of 2 μM

and 4 µM, respectively. The mixture was centrifuged at 25 °C for 30 min at 12,000 x g in a tabletop centrifuge and placed on ice.

Three microliters of the h20S-PA26$^{E102A-Opt5}$ solution were applied to R2/2 400-mesh grids (Quantifoil) that had been plasma treated for 25 s using a glow discharger (Electron Microscopy Sciences) operated under atmospheric gases doped with amylamine. The grids were manually blotted to near dryness with Whatman no. 1 filter paper inside a cold room (4°C, 95% humidity). The sample application and blotting process was repeated twice more to increase on-grid protein concentration. After the third blot, the grid was gravity plunged into liquid ethane using a home-made system and stored under liquid nitrogen.

Cryo-electron microscopy data collection and image processing. Cryo-EM data were acquired using the Leginon software for automated data acquisition[54] using a Titan Krios (Thermo Fisher) equipped with a K2 Summit (Gatan) direct electron detector in counting mode (**Table S2.2**). Movies were collected by navigating to the center of a hole and image shifting a beam of ~600 nm diameter to 10 targets situated at the periphery of the 2-µm hole. A total of 13,329 movies were recorded at a nominal magnification of 29,000 × (1.03 Å magnified pixel size) and composed of 29 frames (250 ms per frame, ~50 e$^-$/Å^2 per movie). Movie collection was guided by real-time assessment of image quality using the Appion image processing environment[55]. Frame alignment and dose weighting were performed in real-time using UCSF Motioncor2[56]. CTF estimation on aligned, unweighted, micrographs was performed with Gctf[57]. All data post-processing steps were conducted in RELION 2.1[52,53].

Single particle analysis was performed in RELION 2.1. Particle picking was conducted using the AutoPick function and resulted in 2,650,143 particle picks that were extracted using a box size of 420 × 420 pixels, down-sampled to 70 × 70 pixels, for reference-free 2D classification. 497,630 particles belonging to the 2D classes demonstrating features characteristic of secondary structural elements were subjected to 3D refinement and subsequent 3D classification (k = 8). A 3D template of the PA26-bound archaeal proteasome (PDB ID: 1YAU) was lowpass-filtered to 60 Å and used to guide the initial 3D refinement and 3D classification. 288,515 particles corresponding to 3D classes without artefactual features were chosen for further data processing. To minimize the detrimental effects of pseudo-symmetry (C2) on resolution, the raw particles were C2 symmetry expanded, 3D refined, and a python script was used to determine the x- and y-shifts required to reposition the proteasome gate at the center of the particle box. The particles at the new center were re-extracted to serve as individual asymmetric units without down-sampling in a box of 256 × 256 pixels, resulting in a total of 577,030 particles. A round of reference-free 2D classification enabled us to remove the ends of *h*20S particles that lacked an activator molecule. This combined expansion and classification approach yielded 521,860 particles. We subjected these particles to 3D classification (k = 8) and selected 384,939 particles whose parent 3D classes resembled a fully assembled *h*20S-PA26[E102A-Opt5] complex. Following removal of particle duplicates with a python script, 326,676 particles were CTF and beam tilt refined. Further 3D classification without alignment (k = 10) revealed two classes with an unresolved alpha subunit helix. We attributed this finding to rotational misalignment around the C7 pseudo-symmetry, longitudinal axis of the complex and discarded these particles. The remaining

247,362 particles were then sorted and pruned by z-score, yielding 234,960 particles for final 3D refinement (2.9 Å) (**Figures S2.6 and S2.7**).

Atomic model building. The atomic model was built by using the *h*20S and PA26 (PDB IDs: 5A0Q and 1YAU, respectively) as templates and rigid body fitting each subunit with Chimera into the electron density map. Subunits for which there was no density as a result of the symmetry expansion and re-extraction processing approach, were removed from the template. The modified templates were subject to one cycle of morphing and simulated annealing in PHENIX, followed by a total of 5 real-space refinement macrocycles with atomic displacement parameters, secondary structure restraints, local grid searches, and global minimization. After PHENIX refinement, manual real-space refinement was performed in Coot. Multiple rounds of real-space refinement in PHENIX (five macro cycles, Ramachandran and rotamer restraints, no morphing, no simulated annealing) and Coot were performed to address geometric and steric discrepancies identified by the RCSB PDB validation server. To ensure atomic models were not overfit as a result of real-space refinement, map-to-model FSCs were calculated with PHENIX[58] (**Figure S2.7B**). All images were generated using UCSF Chimera and ChimeraX.

Data availability. Structural data are available in the Electron Microscopy Databank and the RCSB Protein Databank (EMDB ID: 22259 and PDB ID: 6XMJ). Additional data supporting the findings of this manuscript are available from the corresponding author upon reasonable request.

References

1. Kwon, Y. T. & Ciechanover, A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem. Sci.* **42**, 873–886 (2017).

2. Groll, M. *et al.* A gated channel into the proteasome core particle. *Nat. Struct. Biol.* **7**, 1062–1067 (2000).

3. Zhang, Z. *et al.* Identification of an activation region in the proteasome activator REGalpha. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2807–2811 (1998).

4. Whitby, F. G. *et al.* Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408**, 115–120 (2000).

5. Köhler, A. *et al.* The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol. Cell* **7**, 1143–1152 (2001).

6. Löwe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution. *Science (80-.).* **268**, 533–539 (1995).

7. Förster, A., Whitby, F. G. & Hill, C. P. The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *EMBO J.* **22**, 4356–4364 (2003).

8. Förster, A., Masters, E. I., Whitby, F. G., Robinson, H. & Hill, C. P. The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol. Cell* **18**, 589–599 (2005).

9. Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's α Ring Opens the Gate for Substrate Entry. *Mol. Cell* **27**, 731–744 (2007).

10. Sadre-Bazzaz, K., Whitby, F. G., Robinson, H., Formosa, T. & Hill, C. P. Structure of

a Blm10 Complex Reveals Common Mechanisms for Proteasome Binding and Gate Opening. *Mol. Cell* **37**, 728–735 (2010).

11. Majumder, P. *et al.* Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 534–539 (2019).

12. Rabl, J. *et al.* Mechanism of Gate Opening in the 20S Proteasome by the Proteasomal ATPases. *Mol. Cell* **30**, 360–368 (2008).

13. Yu, Y. *et al.* Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions. *EMBO J.* **29**, 692–702 (2010).

14. Tian, G. *et al.* An asymmetric interface between the regulatory and core particles of the proteasome. *Nat. Struct. Mol. Biol.* **18**, 1259–1267 (2011).

15. Wehmer, M. *et al.* Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc. Natl. Acad. Sci.* **114**, 1305–1310 (2017).

16. Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12991–12996 (2016).

17. Zhu, Y. *et al.* Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nat. Commun.* **9**, 1–12 (2018).

18. Adams, G. M., Crotchett, B., Slaughter, C. A., DeMartino, G. N. & Gogol, E. P. Formation of proteasome-PA700 complexes directly correlates with activation of peptidase activity. *Biochemistry* **37**, 12927–12932 (1998).

19. Eisele, M. R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell Rep.* **24**, 1301-1315.e5

(2018).

20. De la Peña, A. H., Goodall, E. A., Gates, S. N., Lander, G. C. & Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. *Science (80-.).* **362**, (2018).

21. Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. a. & DeMartino, G. N. Differential roles of the COOH termini of AAA subunits of PA700 (19 S regulator) in asymmetric assembly and Activation of the 26 S proteasome. *J. Biol. Chem.* **283**, 31813–31822 (2008).

22. Kumar, B., Kim, Y. C. & DeMartino, G. N. The C terminus of Rpt3, an ATPase subunit of PA700 (19 S) regulatory complex, is essential for 26 S proteasome assembly but not for activation. *J. Biol. Chem.* **285**, 39523–39535 (2010).

23. Opoku-Nsiah, K. A. & Gestwicki, J. E. Aim for the core: suitability of the ubiquitin-independent 20S proteasome as a drug target in neurodegeneration. *Translational Research* **198**, 48–57 (2018).

24. Trader, D. J., Simanski, S., Dickson, P. & Kodadek, T. Establishment of a suite of assays that support the discovery of proteasome stimulators. *Biochim. Biophys. Acta* **1861**, 892—899 (2017).

25. Gizyńska, M. *et al.* Proline- and Arginine-Rich Peptides as Flexible Allosteric Modulators of Human Proteasome Activity. *J. Med. Chem.* **62**, 359–370 (2019).

26. Glickman, M. H., Rubin, D. M., Fried, V. A. & Finley, D. The Regulatory Particle of the *Saccharomyces cerevisiae* Proteasome. *Mol. Cell. Biol.* **18**, 3149–3162 (1998).

27. Stadtmueller, B. M. *et al.* Structure of a Proteasome Pba1-Pba2 Complex Implications for Proteasome Assembly, Activation, and Biological Function. *J. Biol.*

Chem. **287**, 37371–37382 (2012).

28. Ganesh, A. N., Donders, E. N., Shoichet, B. K. & Shoichet, M. S. Colloidal aggregation: From screening nuisance to formulation nuance. *Nano Today* **19**, 188–200 (2018).

29. Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* **12**, 6–15 (2012).

30. Stadtmueller, B. M. *et al.* Structural models for interactions between the 20S proteasome and its PAN/19S activators. *J. Biol. Chem.* **285**, 13–17 (2010).

31. Smith, D. M. *et al.* ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. *Mol. Cell* **20**, 687–698 (2005).

32. Yu, Z. *et al.* Allosteric coupling between α-rings of the 20S proteasome. *Nat. Commun.* 1–11 (2019). doi:10.1101/832113

33. Thibaudeau, T. A., Anderson, R. T. & Smith, D. M. A common mechanism of proteasome impairment by neurodegenerative disease-associated oligomers. *Nat. Commun.* **9**, (2018).

34. Eisele, M. R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell Rep.* **24**, 1301-1315.e5 (2018).

35. Harshbarger, W., Miller, C., Diedrich, C. & Sacchettini, J. Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib. *Structure* **23**, 418–424 (2015).

36. Park, S. *et al.* Reconfiguration of the proteasome during chaperone-mediated assembly. *Nature* **497**, 512–6 (2013).

37. Huang, X., Luan, B., Wu, J. & Shi, Y. An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* **23**, 778–785 (2016).

38. Chelli, R., Gervasio, F. L., Procacci, P. & Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J. Am. Chem. Soc.* **124**, 6133–6143 (2002).

39. Knowlton, J. R. *et al.* Structure of the proteasome activator REGalpha (PA28alpha). *Nature* **390**, 639–643 (1997).

40. Barthelme, D. & Sauer, R. T. Bipartite determinants mediate an evolutionarily conserved interaction between Cdc48 and the 20S peptidase. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3327–32 (2013).

41. Kiessling, L. L., Gestwicki, J. E. & Strong, L. E. Synthetic multivalent ligands as probes of signal transduction. *Angew. Chemie - Int. Ed.* **45**, 2348–2368 (2006).

42. Toste Rêgo, A. & da Fonseca, P. C. A. Characterization of Fully Recombinant Human 20S and 20S-PA200 Proteasome Complexes. *Mol. Cell* **76**, 138-147 (2019).

43. Li, X., Thompson, D., Kumar, B. & DeMartino, G. N. Molecular and cellular roles of PI31 (PSMF1) protein in regulation of proteasome function. *J. Biol. Chem.* **289**, 17392–17405 (2014).

44. Ramachandran, K. V & Margolis, S. S. A mammalian nervous-system-specific plasma membrane proteasome complex that modulates neuronal function. *Nat. Struct. Mol. Biol.* **24**, 419–430 (2017).

45. Hirano, Y. *et al.* A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. *Nature* **437**, 1381–1385 (2005).

46. Schmidt, M. & Finley, D. Regulation of proteasome activity in health and disease.

Biochim. Biophys. Acta **1843**, 13–25 (2014).

47. Li, J., Powell, S. R. & Wang, X. Enhancement of proteasome function by

    PA28 overexpression protects against oxidative stress. *FASEB J.* **25**, 883–893

    (2011).

48. Qu, J. *et al.* Specific Knockdown of α-Synuclein by Peptide-Directed Proteasome

    Degradation Rescued Its Associated Neurotoxicity. *Cell Chem. Biol.* (2020).

    doi:https://doi.org/10.1016/j.chembiol.2020.03.010

49. Jones, C. L., Njomen, E., Sjögren, B., Dexheimer, T. S. & Tepe, J. J. Small Molecule

    Enhancement of 20S Proteasome Activity Targets Intrinsically Disordered Proteins.

    ACS Chem. Biol. **12**, 2240–2247 (2017).

50. Coleman, R. A. & Trader, D. J. Development and Application of a Sensitive Peptide

    Reporter to Discover 20S Proteasome Stimulators. *ACS Comb. Sci.*

    acscombsci.7b00193 (2018). doi:10.1021/acscombsci.7b00193

51. Crooks, G., Hon, G., Chandonia, J. & Brenner, S. WebLogo: A Sequence Logo

    Generator. *Genome Res* **14**, 1188–1190 (2004).

52. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure

    determination. *Nat. Methods* **9**, 853–854 (2012).

53. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-EM

    structure determination with parallelisation using GPUS in RELION-2. *Elife* **5**, 1–21

    (2016).

54. Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *J.*

    Struct. Biol. **151**, 41–60 (2005).

55. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM

image processing. *J. Struct. Biol.* **166**, 95–102 (2009).

56. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

57. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

58. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
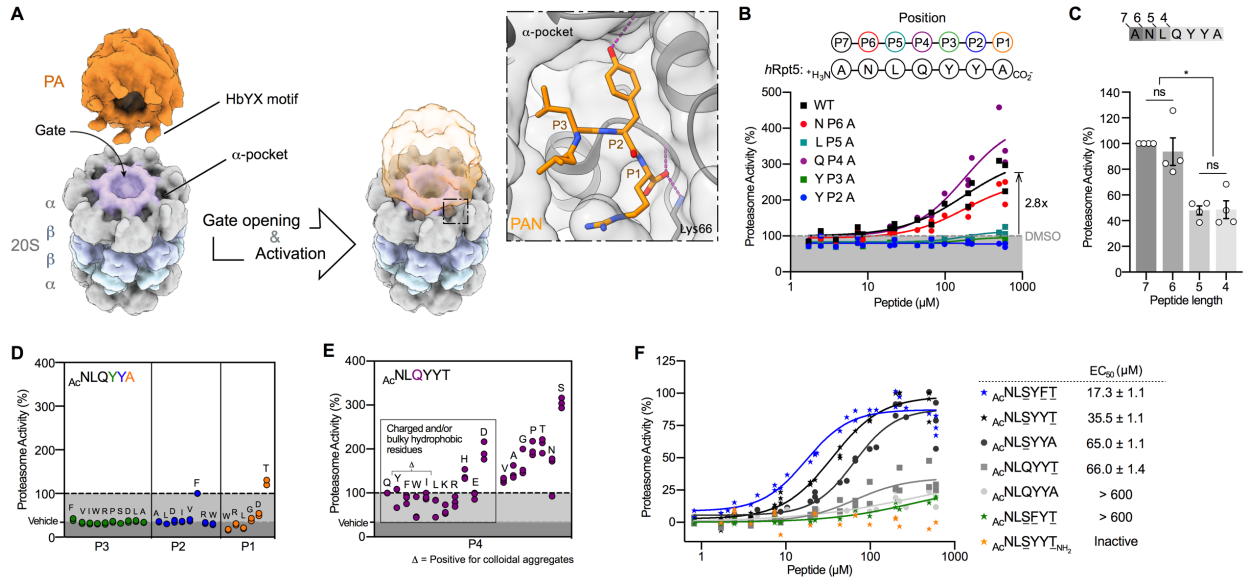
Figure 2.1. Activation of *h*20S by *h*Rpt5-derived peptides reveals unexpected SAR.
A, Schematic of proteasome activation by the C-terminal HbYX motif of a model PA. HbYX motifs are known to dock into α-pockets located between subunits of the 20S α-rings (gray). This interaction opens the gates of the 20S, allowing substrate entry. The inset summarizes the interactions between PAN's HbYX motif (orange) and the α-pocket (gray surface/cartoon) of the archaeal 20S (PDB ID: 3IPM). **B**, Alanine scanning revealed residues of Rpt5 C-terminus that make key contributions to stimulation of the *h*20S. Activation of the *h*20S (4 nM) was measured by an increase in the hydrolysis rate (RFU/s) of the fluorescent substrate, suc-LLVY-amc (10 μM) upon addition of *h*Rpt5-derived peptides. Data are normalized to DMSO-treated *h*20S and plotted individually (n = 3). **C**, Proteasome stimulation by peptides (250 μM) from an N-terminal truncation series. Results are normalized to ANLQYYA and the average of four independent experiments (open circles; n = 4) is plotted with error reported as s.e.m. P-values were calculated using two-tailed unpaired t-test: *p < 0.04. **D**, Scatterplot of the relative activities of N-terminally acetylated *h*Rpt5 hexapeptides (250 μM) sampling modifications along the P3 (green), P2 (blue) and P1 (orange) residues. Data are normalized to Ac-NLQYYA and plotted individually (n = 2). **E**, Scatterplot of maximal activity of hexapeptides sampling all proteinogenic amino acids at the P4 position (purple). Data are normalized to Ac-NLQYYT and the average of three independent experiments (circles; n = 3) is plotted with error reported as s.e.m. **F**, Combination of optimized residues (underlined) from the SAR campaign. Data are normalized to Opt5 (Ac-NLSYYT, star) and plotted individually (n = 3 or 4). EC_{50} values were calculated from four independent experiments and reported as mean ± s.e.m.
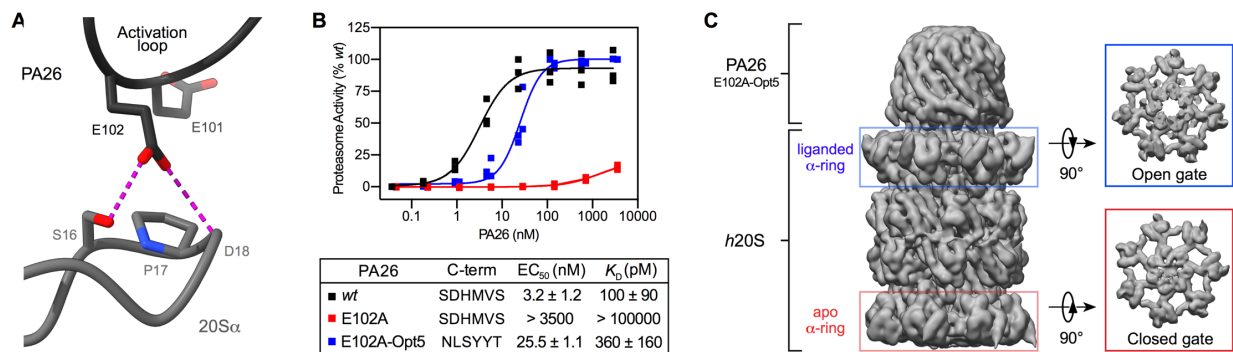
Figure 2.2. PA26$^{E102A-Opt5}$ induces terminus-dependent activation and gate opening of the *h*20S.

A, Schematic of how PA26's activation loop (black cartoon) directly repositions the Pro17 reverse turn of the 20S α-subunit (gray cartoon) to induce HbYX-independent gate opening (PDB ID: 1YA7). **B**, Stimulation of *h*20S by PA26 constructs: wild type PA26 (*wt*, black), disabled-loop mutant (PA26^{E102A}, red), and a disabled-loop mutant with C-terminal Opt5 (NLSYYT) sequence (PA26$^{E102A-Opt5}$, blue). Data are normalized to PA26$^{E102A-Opt5}$ and plotted individually (n = 3). Reported EC$_{50}$ is a mean of EC$_{50}$ values calculated from three independent experiments (n = 3) with error reported as s.e.m. K_D values were determined by BLI from three independent experiments (n = 3) and reported as the mean ± s.e.m. **C**, PA26$^{E102A-Opt5}$ induces gate opening. The 3D reconstruction of the singly capped PA26$^{E102A-Opt5}$-*h*20S complex at 2.9 Å resolution from 234,960 particles shows gate opening (blue box) relative to the apo, closed-gate α-ring on the opposite side (red box).
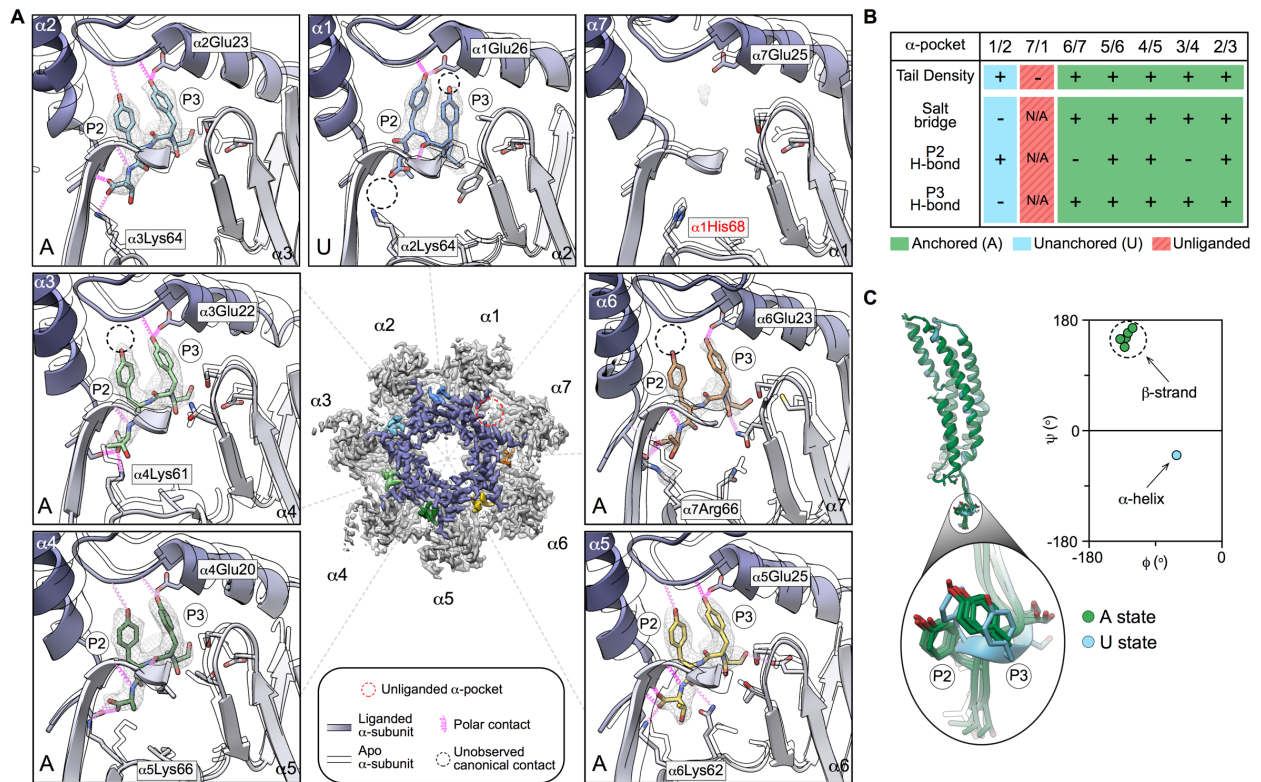
Figure 2.3. Ensemble of C-terminal interactions promote gate opening in the *h*20S. **A**, Opt5 sequence in the *h*20S α-pockets. The α-subunits are numbered according to the yeast proteasome. *Center*, a top-view 3D density map of the liganded α-ring displaying color-coded densities for each C-tail. The empty α-pocket (α7/α1) is marked by a red dotted circle. Close-up views of each C-terminal Opt5 sequence (stick) is superimposed over their corresponding cryo-EM densities (mesh). PA26[E102A-Opt5] induces terminus-dependent conformational changes throughout the inner (purple) and outer (gray) regions of the α-pockets relative to the apo *h*20S (outlined) (PDB ID: 4R3O). Predicted interactions (pink coils) and unobserved canonical interactions (black dotted circles) are denoted. Anchored (A) or unanchored (U) Opt5 binding states are labeled at the bottom-left corner of each panel (see text). **B**, Summary table of the interactions for each α-pocket. **c**, Opt5 binds with distinct conformations in the A and U states. Opt5 forms an α-helix In the U state (α1/α2-pocket), while it adopts a β-strand conformation in the A state. Overlaying the C-tails highlights the distinct backbone conformations between the A and U states, while also showing the striking similarities between the A states, especially in the position of the P2 and P3 Tyr residues. See **Table S2.3** for relative densities of the C-tails (**B**) and torsion angles (**C**).
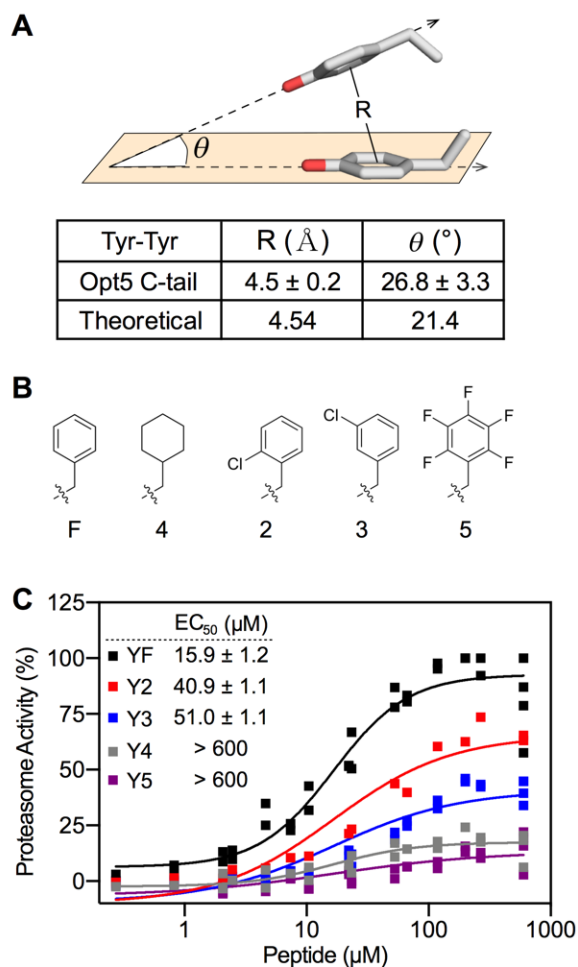
Figure 2.4. P3 and P2 residues make intramolecular π-stacking interactions within the *h*20S α-pockets.

A, Experimental and theoretical distances (*R*) and angles of incidence (*θ*) for adjacent P2 and P3 tyrosine residues[38]. Values are of bound Opt5 from the cryo-EM structure (n = 6) and reported as mean ± s.d. See **Table S2.3** for individual measurements. **B,** Structure of the side chains from unnatural amino acids used in this study. **C,** P2 modifications to the π electron density decreases the stimulatory effect of Opt5. Data are normalized to Opt5^{YF} and plotted individually (n = 2 to 4). Reported EC$_{50}$ is a mean of EC$_{50}$ values calculated from two to four independent experiments with error reported as s.e.m.
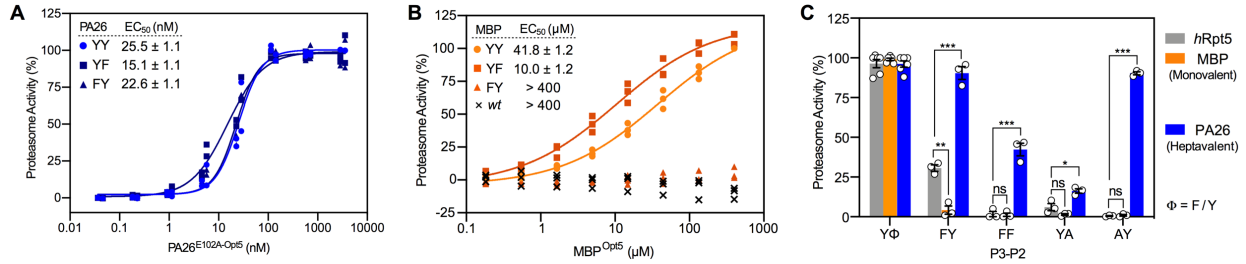
Figure 2.5. Valency tunes the sequence preferences of *h*20S PAs.
A, Proteasome activity of heptameric PA26[E102A-Opt5] (YY, circle) and C-terminally modified variants, YP2F mutant (YF, square) and YP3F mutant (FY, triangle). **B**, Similar to **A** except using C-terminally modified mutants of MBP, tested alongside wildtype MBP (*wt*, 'x'). Data are normalized to YY and plotted individually (n = 3 or 4). Reported EC_{50} is a mean of EC_{50} values calculated from three independent experiments (n = 3 or 4) with error reported as s.e.m. **C,** Proteasome stimulation by monovalent Opt5 peptides (gray), monovalent MBP[Opt5] chimeras (orange) and multivalent PA26[E102A-Opt5] (blue). YΦ is a composite of YY and YF activators. Data are calculated from three independent experiments (see **Figure S2.4B-D**), normalized to YY, and plotted individually (open circles, n = 6 or 3) with error reported as s.e.m. P-values were calculated using unpaired t-test: *p < 0.05, **p < 0.01, and ***p < 0.001.
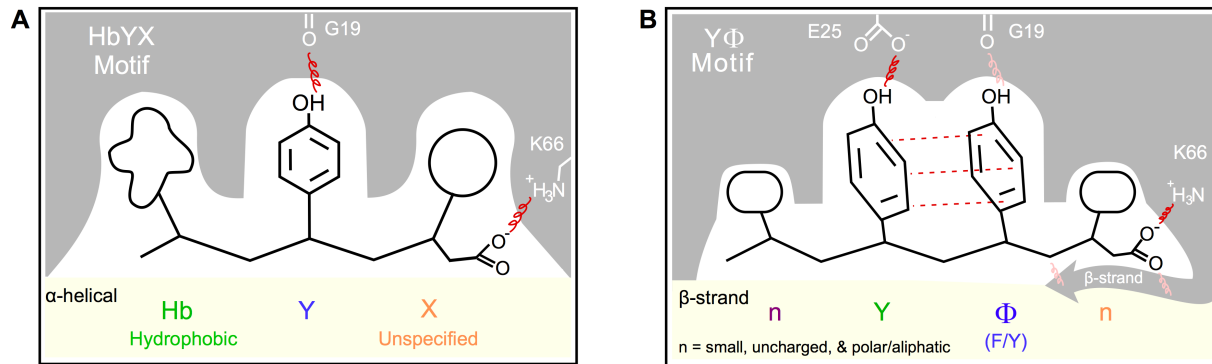
Figure 2.6. The YΦ motif re-defines the model for terminus-dependent *h*20S activation.
A, Model summarizing the HbYX motif. **B**, Redefined model of the nYΦn (YΦ) motif summarizes the proposed mechanism for termini-mediated, monovalent activators of the human proteasome. Interactions denoted for the YΦ motif seem specific to monovalent activators of the *h*20S. The critical interactions (red), including polar contacts (squiggle) and potential π stacking (dotted line), and less critical interactions (pink) are denoted for the P1 (orange), P2 (blue), P3 (green) and P4 (purple) residues of the motif. The predicted binding conformation and residue preferences of the motif are also described.
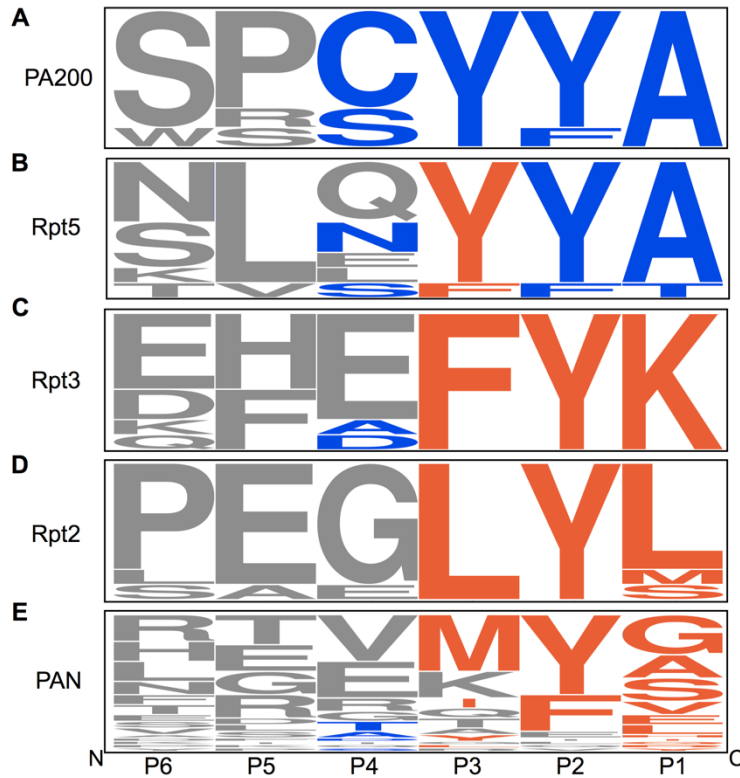
Figure 2.7. The YΦ motif is strictly conserved in the eukaryotic, monovalent activator PA200.

A-E, WebLogo (http://weblogo.berkeley.edu)[51] representation of the last six residues of (**A**) PAN (n = 41), (**B**) PA200 (n = 7)**,** (**C**) Rpt5 (n = 9), (**D**) Rpt3 (n = 9), and (**E**) Rpt2 (n = 9). Amino acid residues at the P1, P2, P3 and P4 (YΦ only) positions are colored according to sequence preferences of the HbYX (orange) or nYΦn (blue) motifs. Unspecified, along with P5 and P6 residues are in gray. The height of each residue represents the relative frequency. The complete entry list in **Table S2.4**.
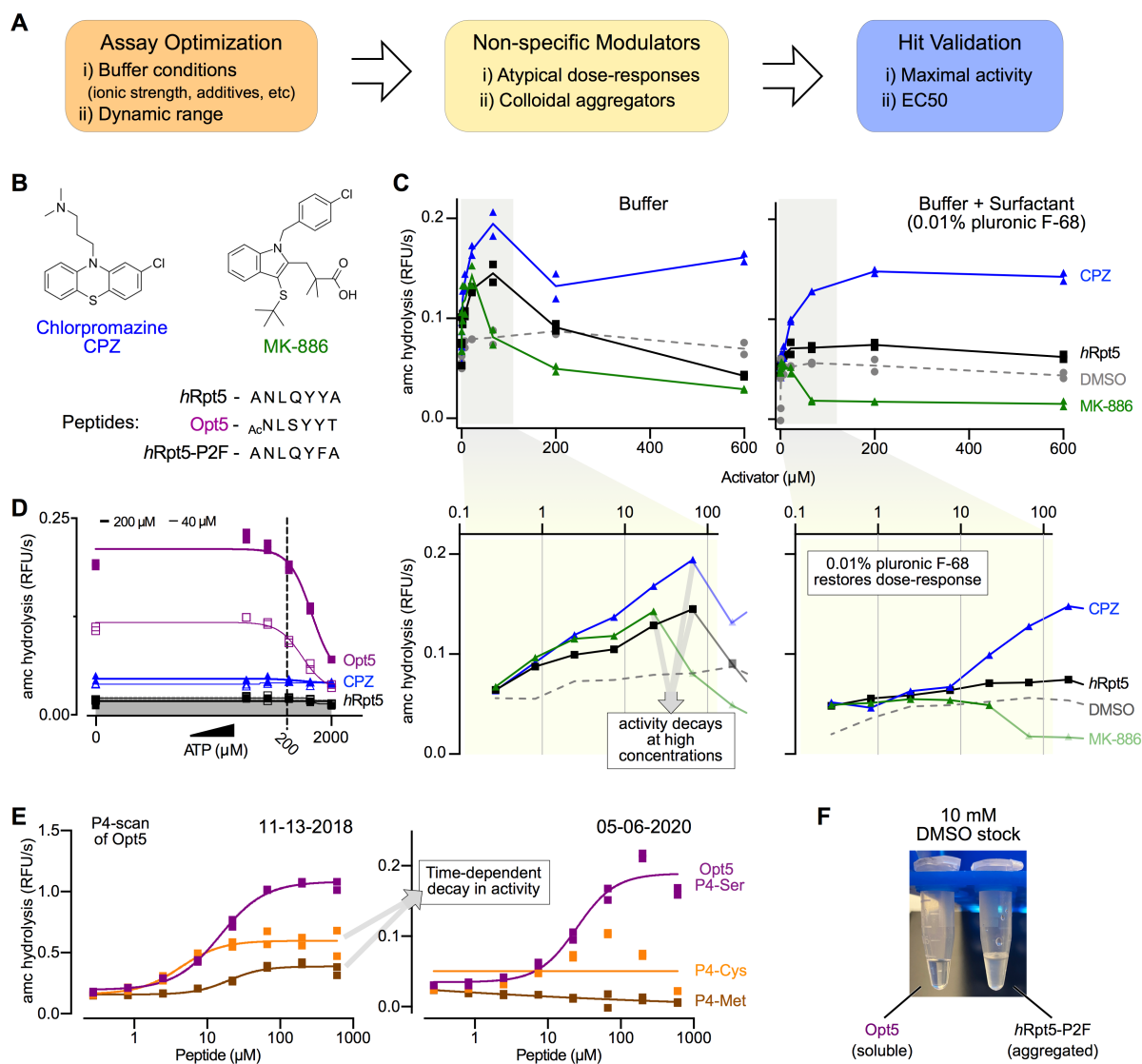
Figure S2.1. Assay optimization and hit validation.
A, Pipeline for effective discovery of peptide activators of the *h*20S. **B**, Structures of previously reported 20S activators and peptide sequences used in this study. **C**, Addition of surfactant (0.01% Pluronic F-68 ®) restores dose-dependence of *h*20S activation. Representative dose-response curves for small molecule (CPZ, blue; MK-886, green) and *h*Rpt5 peptide (black) activators with zoom-in of non-specific inhibition at higher concentrations. **D**, ATP inhibits activation by *h*Rpt5-based peptides. Representative dose-response curve of activation with ATP titration. Assay buffer will contain 200 μM of ATP. **E**, Time-dependent loss of activity in Cys- and Met-containing peptides. Two separate dose-response curves from P4-scan (see **Figure S2.4B**). **C**-**E**, Technical replicates are plotted individually (n = 2). **F,** Typical, soluble peptide appears clear as 10 mM stock in DMSO (left, Opt5), while aggregated peptide appears white and cloudy (right, *h*Rpt5-P2F).
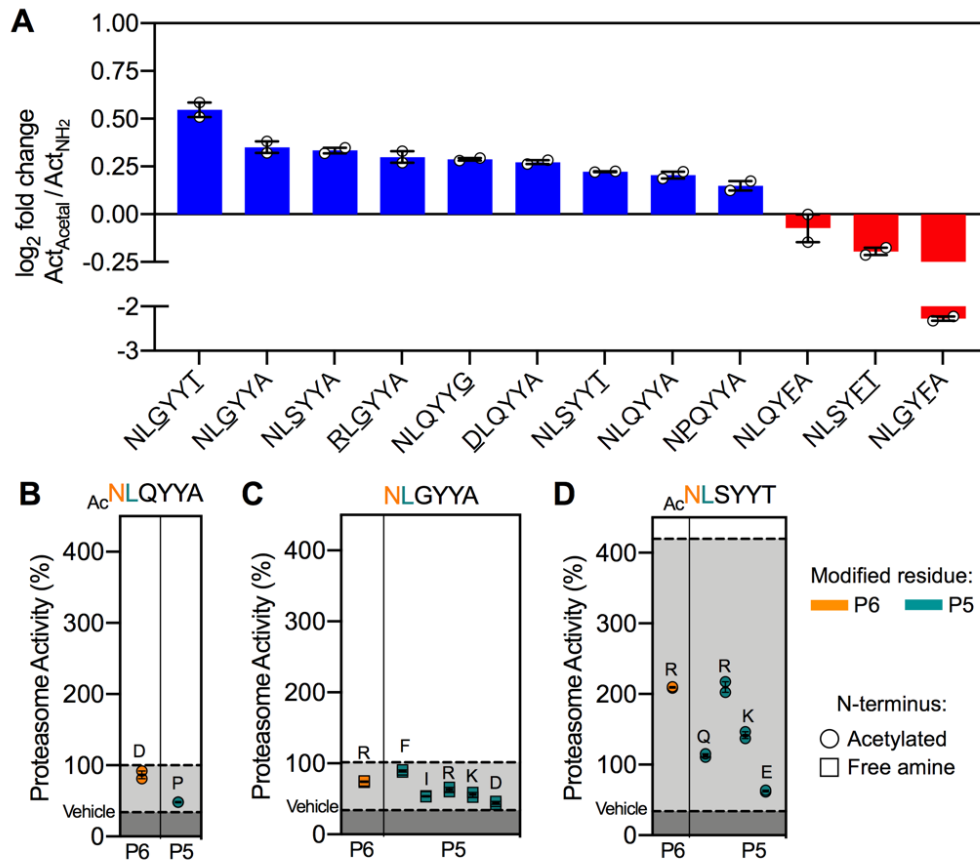
Figure S2.2. Effects of N-terminal modifications on stimulation of *h*20S by peptides. **A**, N-terminal acetylation of *h*Rpt5-derived peptides tends to modestly enhance their stimulatory activity. The effect of N-terminal acetylation is plotted as $\log_2$ fold increase (blue) or decrease (red) in proteasomal stimulation. Additional modifications from the wild type *h*Rpt5 sequence are underlined. Reported data is a mean calculated from two independent experiments with error reported as s.e.m. **B-D**, Scatterplots of the relative activities of peptides (250 μM) sampling residues at the P6 (orange) and P5 (teal) positions. Activities are benchmarked by vehicle and either (**B**) Ac-NLQYYA, (**C**) NLGYYA, or (**D**) Ac-NLSYYT (dotted lines). Data are normalized to Ac-NLQYYA and plotted individually (n = 2) along with the mean ± s.e.m.
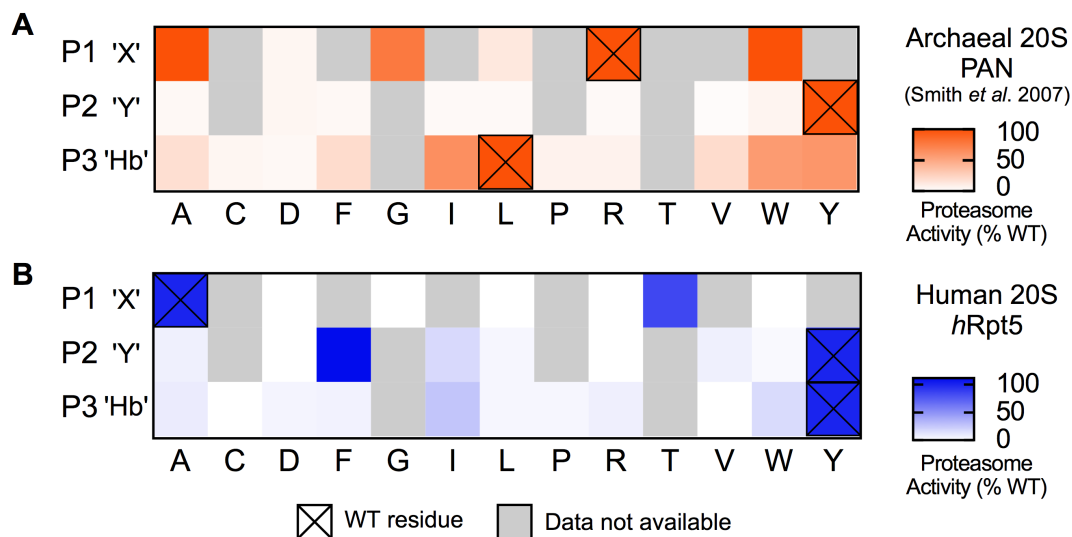
Figure S2.3. The structure-activity profile for *h*Rpt5-activated *h*20S differs from the HbYX model.
A, Summary of the previously described SAR for PAN-stimulated activity of the Thermoplasma 20S proteasome[9] (orange). **B**, The SAR for *h*Rpt5-stimulated activity of the human 20S (blue). The corresponding mutations from Smith et al. 2007 were made to Ac-NLQYYA and the data aligned to the archaeal heatmaps. Data are normalized to WT and plotted as mean (n = 2). **A**,**B**, Residues of wildtype sequences are denoted with an 'x'. Activity measurements were not determined for residues in gray.
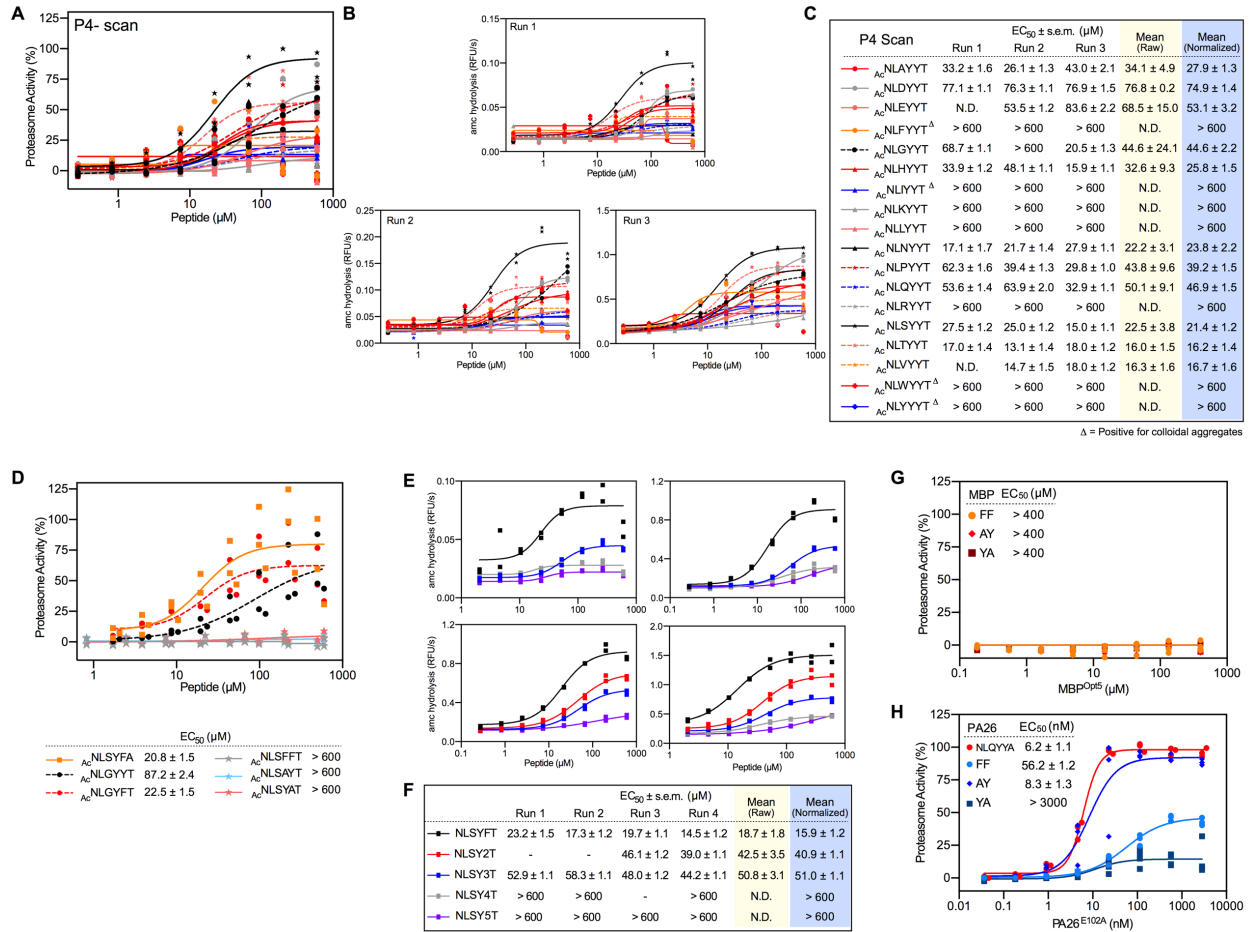
Figure S2.4. Dose-dependent stimulation induced by different PAs.
A-C, Stimulation of *h*20S by P4-substituted hexapeptides. (**A**) Normalized dose-response curves. Data are normalized to Opt5 (Ac-NLSYYT) and plotted individually (n = 3). (**B**) Experimental replicates of non-normalized dose-response curves with technical replicate plotted individually (n=2). (**C**) EC_{50} values are reported for each independent experiment (n = 3). The mean of the "raw" EC_{50} values (yellow) along with the mean of the normalized EC_{50} values (blue) are reported with error as s.e.m. (n = 3). **D**. Dose-response curves for additional *h*Rpt5-derived hexapeptides. Data are normalized to Opt5 and plotted individually (n = 3) with calculated EC_{50} values with error reported as s.e.m. (n = 3). **E,F**, Activity of hexapeptides for pi-stacking studies. (**E**) Experimental replicates of raw dose-response curves with technical replicate plotted individually (n=2). (**F**) EC_{50} values are reported for each independent experiment (n = 4). The mean of the "raw" EC_{50} values (yellow) along with the mean of the normalized EC_{50} values (blue) are reported with error as s.e.m. (n = 2 to 4). **G,H**, Dose-response curves for additional (**G**) C-terminal MBP mutants and (**H**) C-terminal PA26[E102A] mutants. Data are normalized to Opt5 sequence and plotted individually (n = 3). Reported EC_{50} is a mean of EC_{50} values calculated from two to four independent experiments with error reported as s.e.m.

Panel C — P4 Scan

P4 Scan	Run 1	Run 2	Run 3	Mean (Raw)	Mean (Normalized)
Ac-NLAYYT	33.2 ± 1.6	26.1 ± 1.3	43.0 ± 2.1	34.1 ± 4.9	27.9 ± 1.3
Ac-NLDYYT	77.1 ± 1.1	76.3 ± 1.1	76.9 ± 1.5	76.8 ± 0.2	74.9 ± 1.4
Ac-NLEYYT	N.D.	53.5 ± 1.2	83.6 ± 2.2	68.5 ± 15.0	53.1 ± 3.2
Ac-NLFYYT Δ	> 600	> 600	> 600	N.D.	> 600
Ac-NLGYYT	68.7 ± 1.1	> 600	20.5 ± 1.3	44.6 ± 24.1	44.6 ± 2.2
Ac-NLHYYT	33.9 ± 1.2	48.1 ± 1.1	15.9 ± 1.1	32.6 ± 9.3	25.8 ± 1.5
Ac-NLIYYT Δ	> 600	> 600	> 600	N.D.	> 600
Ac-NLKYYT	> 600	> 600	> 600	N.D.	> 600
Ac-NLLYYT	> 600	> 600	> 600	N.D.	> 600
Ac-NLNYYT	17.1 ± 1.7	21.7 ± 1.4	27.9 ± 1.1	22.2 ± 3.1	23.8 ± 2.2
Ac-NLPYYT	62.3 ± 1.6	39.4 ± 1.3	29.8 ± 1.0	43.8 ± 9.6	39.2 ± 1.5
Ac-NLQYYT	53.6 ± 1.4	63.9 ± 2.0	32.9 ± 1.1	50.1 ± 9.1	46.9 ± 1.5
Ac-NLRYYT	> 600	> 600	> 600	N.D.	> 600
Ac-NLSYYT	27.5 ± 1.2	25.0 ± 1.2	15.0 ± 1.1	22.5 ± 3.8	21.4 ± 1.2
Ac-NLTYYT	17.0 ± 1.4	13.1 ± 1.4	18.0 ± 1.2	16.0 ± 1.5	16.2 ± 1.4
Ac-NLVYYT	N.D.	14.7 ± 1.5	18.0 ± 1.2	16.3 ± 1.6	16.7 ± 1.6
Ac-NLWYYT Δ	> 600	> 600	> 600	N.D.	> 600
Ac-NLYYYT Δ	> 600	> 600	> 600	N.D.	> 600

Δ = Positive for colloidal aggregates

Panel D — EC_{50} (µM)

	EC_{50} (µM)			EC_{50} (µM)
Ac-NLSYFA	20.8 ± 1.5		Ac-NLSFFT	> 600
Ac-NLGYYT	87.2 ± 2.4		Ac-NLSAYT	> 600
Ac-NLGYFT	22.5 ± 1.5		Ac-NLSYAT	> 600

Panel F

	Run 1	Run 2	Run 3	Run 4	Mean (Raw)	Mean (Normalized)
NLSYFT	23.2 ± 1.5	17.3 ± 1.2	19.7 ± 1.1	14.5 ± 1.2	18.7 ± 1.8	15.9 ± 1.2
NLSY2T	-	-	46.1 ± 1.2	39.0 ± 1.1	42.5 ± 3.5	40.9 ± 1.1
NLSY3T	52.9 ± 1.1	58.3 ± 1.1	48.0 ± 1.2	44.2 ± 1.1	50.8 ± 3.1	51.0 ± 1.1
NLSY4T	> 600	> 600	-	> 600	N.D.	> 600
NLSY5T	> 600	> 600	> 600	> 600	N.D.	> 600

Panel G — MBP EC_{50} (µM)

	EC_{50} (µM)
FF	> 400
AY	> 400
YA	> 400

Panel H — PA26 EC_{50} (nM)

	EC_{50} (nM)
NLQYYA	6.2 ± 1.1
FF	56.2 ± 1.2
AY	8.3 ± 1.3
YA	> 3000

Figure S2.5. Distinct proteasome substrates report similar SAR for *h*Rpt5-based activators.
A,B, Stimulation of the *h*20S (4 nM) by *h*Rpt5-based peptides was assessed with (**A**) the trypsin-targeted boc-LRR-amc (20 μM) and (**B**) the gate-sensitive nonapeptide FAM-LFP (100 nM) as alternate substrates to suc-LLVY-amc. Data are normalized to (**A**) Ac-NLSYFT or (**B**) Ac-NLSYYT and plotted individually (n = 2). **C,D** EC$_{50}$ values determined with LRR and LFP substrates were calculated from two independent experiments and reported as the mean ± s.e.m. Corresponding values for LLVY were also included. **D-F**, Experimental replicates of non-normalized dose-response curves for optimization panel of *h*Rpt5 assessing hydrolysis of (**D**) LLVY, (**E**) LRR, and (**F**) LFP substrates. Technical replicates (n=2) are plotted individually for two to four independent experiments with the EC$_{50}$ values ± s.e.m. reported for each replicate. The "raw" mean (yellow) EC$_{50}$ was calculated and reported alongside the "normalized" mean (blue) with error reported as s.e.m. (n = 2 or 4).

78
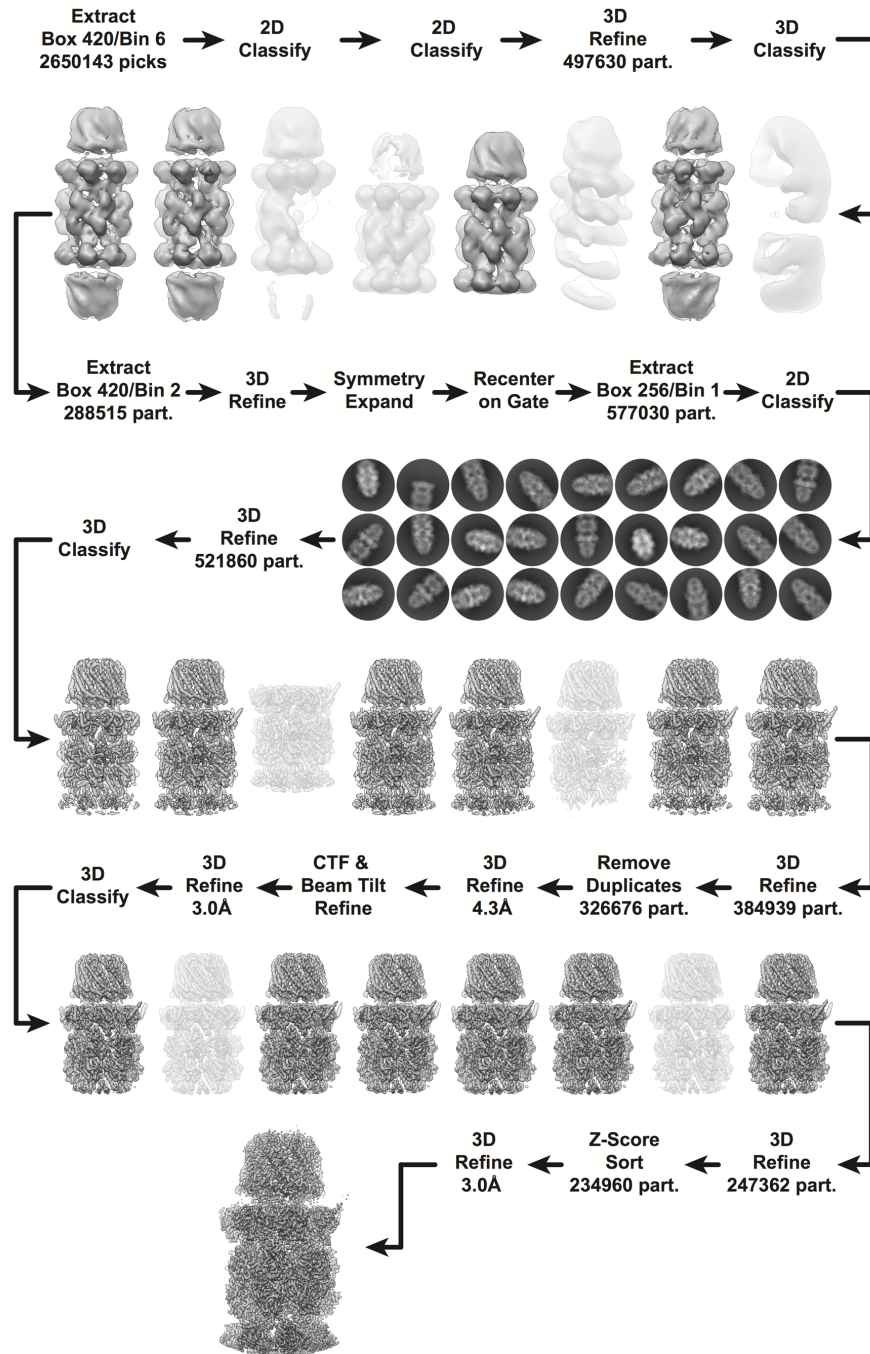
Figure S 2.6. Schematic for cryo-EM single-particle data processing.
Low resolution or artefactual reconstructions (white) and corresponding particles were excluded from subsequent processing steps, whereas all other reconstructions (gray) and corresponding particles were utilized. Representative 2D classes (dark circular background) are shown.
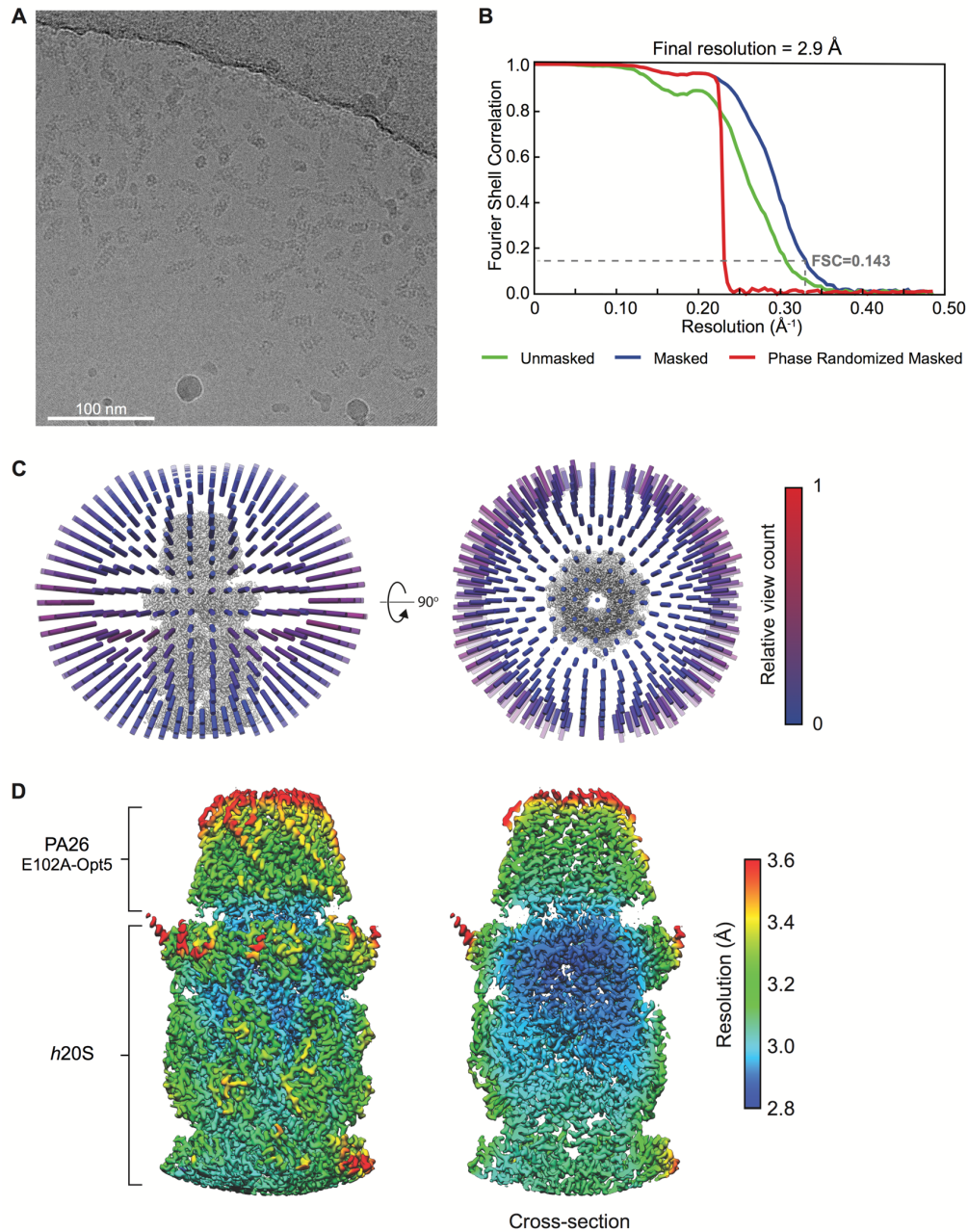
Figure S2.7. Cryo-EM metrics for PA26[E102A-Opt5]**-*h*20S complex.**
A, Representative micrograph. **B**, The Fourier Shell Correlation (FSC) curves for the unmasked (green), masked (blue) and phase randomized (red) reconstructions. The resolution at 0.143 is indicated by a dashed line. **C**, Distribution of Euler angles displaying the orthogonal view of the reconstruction. The angular distribution is shown as a column whose longitudinal axis aligns with the normal of the corresponding back-projection **D**, Local resolution estimates of the global reconstruction for the complete volume (left) and a coronal cross-section (right). **C,D** Generated and calculated with RELION[52,53].
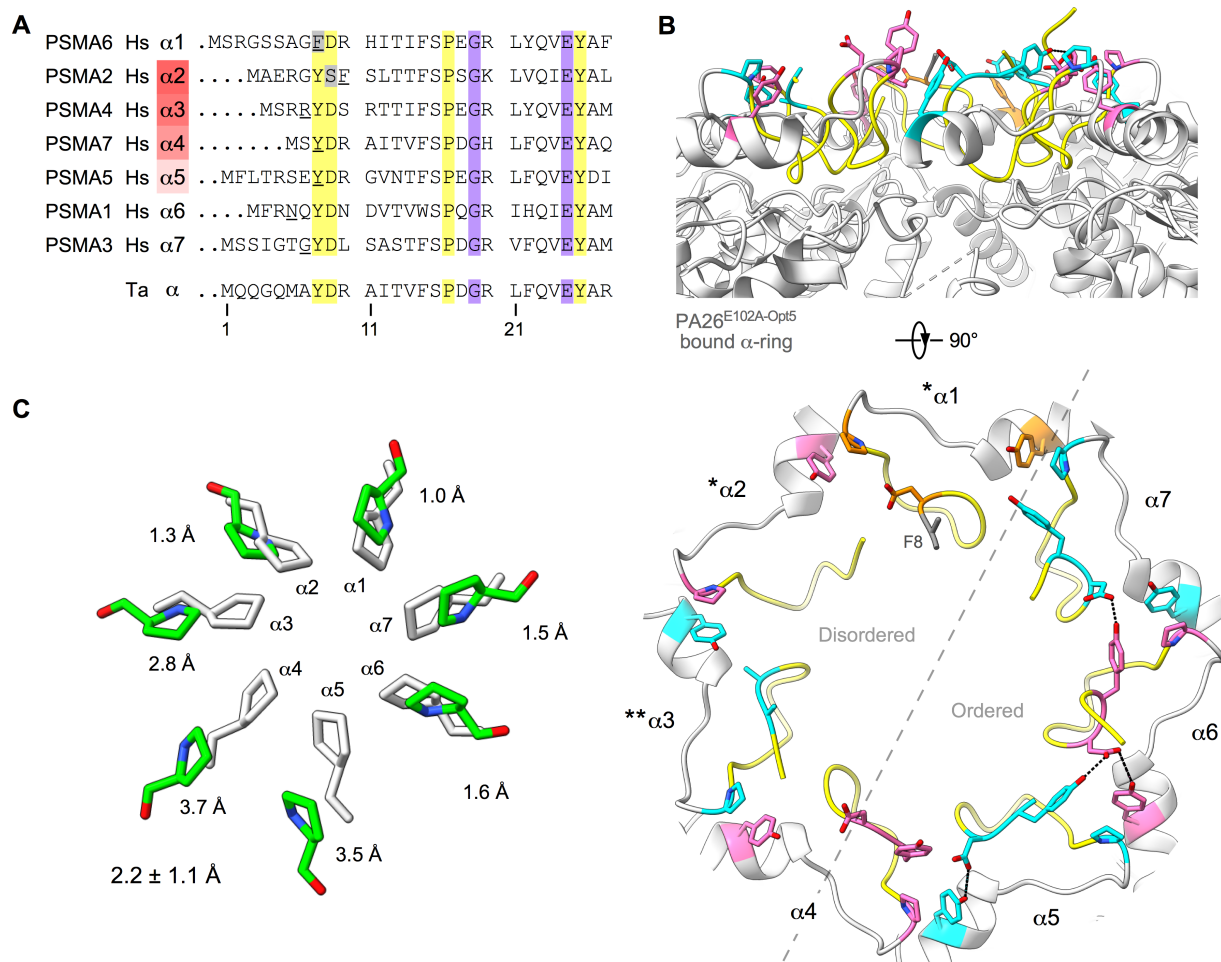
Figure S2.8. Asymmetric *h*20S open-gate conformation.
A, N-terminal sequence alignment of 20S proteasome α-subunits from *H. sapien* (Hs) and *T. acidophilum* (Ta). The α-subunits are labeled with human gene name but numbered according to yeast α-subunit/yeast gene. The α-subunits that directly gate the α-annulus (red) with their relative contributions to gating denoted with a gradient. The most N-terminal residue with resolvable backbone density in the cryo-EM reconstruction is underlined. The conserved cluster (yellow), non-canonical residues (gray), and conserved α-pocket residues (purple) are highlighted. **B**, Conformational state of N-terminal gate (yellow) and interactions of conserved cluster (α1, orange; α2, α4, α6, pink; α3, α5, α7, cyan) that stabilize asymmetric open-gate *h*20S. Non-canonical residues (gray) are denoted for resolved residues. Single or double asterisks indicates α-subunits featuring unresolved extreme N-termini with non-canonical (*) or conserved (**) residues, respectively. **C**, Radial displacement of αPro17. Individual measurements are reported along with mean ± s.d.
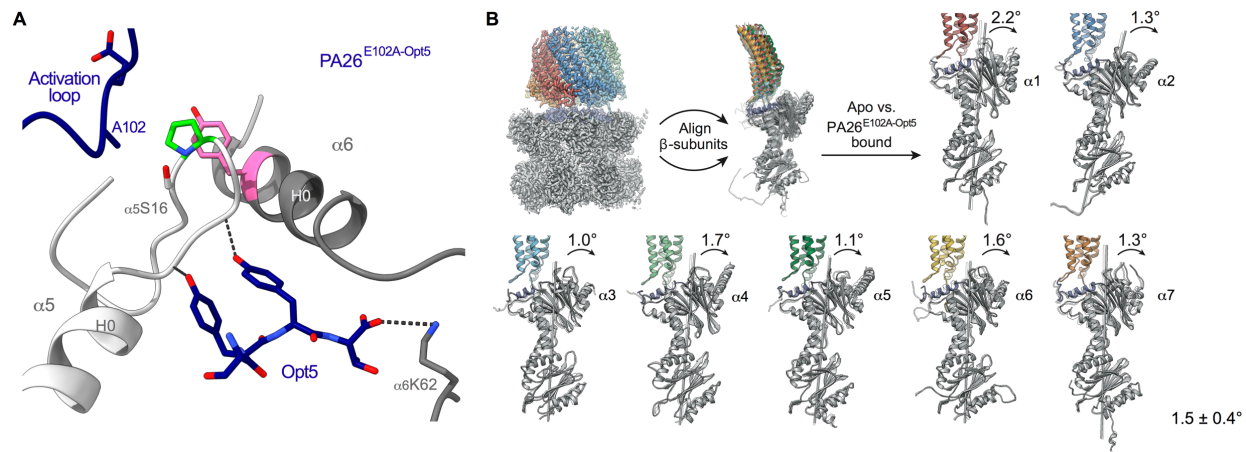
Figure S2.9. Structural evidence implicates termini-mediated gate opening.
A, Disabled activation loop of PA26$^{E102A\text{-}Opt5}$ no longer interacts with αPro17 reverse turn in representative α-pocket (α5/α6-pocket). C-terminal Opt5 engages residues within reverse-turn loop from the α-pocket. **B**, PA26$^{E102A\text{-}Opt5}$ induces rotation of the α-subunits about the axial channel. Individual measurements are reported along with mean ± s.d.

A Rpt5 (PA700)
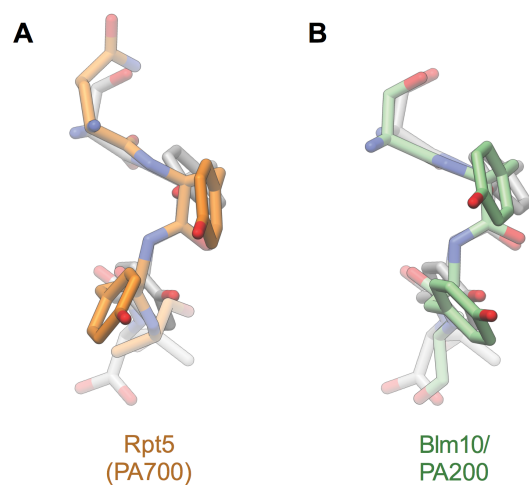
B Blm10/ PA200

Figure S2.10. Binding orientation of reported YΦ motifs align with Opt5.
A,B, Overlay highlights similar backbone conformation between bound Opt5 (SYYT, grey) and the C-terminus of (**A**) the human Rpt5 (QYYA, orange) (PDB ID: 5GJR) or (**B**) the yeast Blm10/PA200 (SYYA, green) (PDB ID: 4V70). Aromatic P2 and P3 Tyr residues share π-stacking orientation. Interactions with α-pocket residues are denoted.
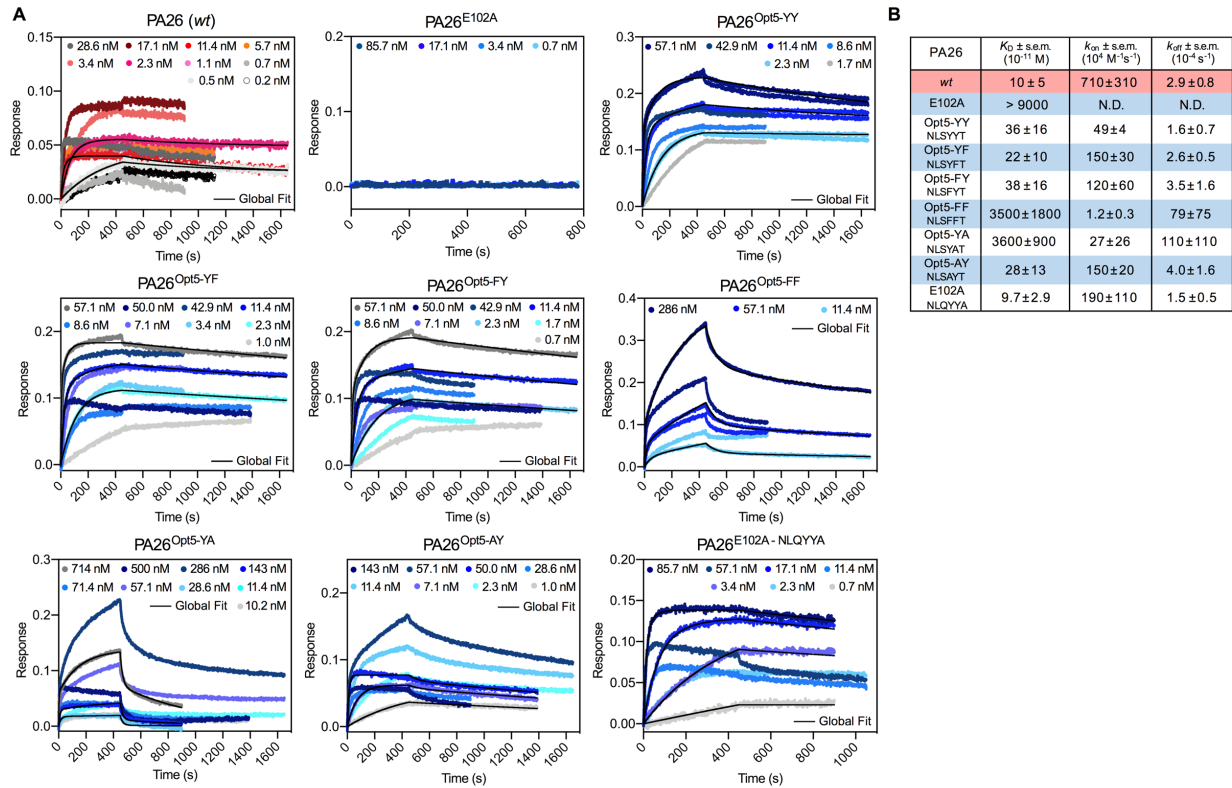
Figure S2.11. Biolayer interferometry (BLI) of PA26 activators.
A, Sensorgram traces of *h*20S binding to either wildtype (*wt*, red) or E102A (blue) variants of PA26 at different concentrations. Colored curves represent measurements for individual experiments (n = 2 or 3) and solid black lines correspond to global-fit for a representative experiment. **B,** Binding and kinetic constants measured by BLI for binding of PA26 to *h*20S. N.D. denotes values that could not be determined due to undetectable binding.

Table S2.1. *h*Rpt5-derived peptides.

ID	Peptide length	N-term Modification	Sequence	Modification
1	7	-	ANLQYYA	-
2	7	-	AALQYYA	NP6A
3	7	-	ANAQYYA	LP5A
4	7	-	ANLAYYA	QP4A
5	7	-	ANLQAYA	YP3A
6	7	-	ANLQYAA	YP2A
7	6	-	NLQYYA	-
8	5	-	LQYYA	-
9	4	-	QYYA	-
10	6	Ac	NLQYYA	-
11	6	Ac	NLQYYA-CONH$_2$	C-term amide
12	6	-	NLQYYD	AP1D
13	6	Ac	NLQYYD	AP1D
14	6	-	NLQYYG	AP1G
15	6	Ac	NLQYYG	AP1G
16	6	Ac	NLQYYL	AP1L
17	6	Ac	NLQYYR	AP1R
18	6	Ac	NLQYYT	AP1T
19	6	Ac	NLQYYW	AP1W
20	6	Ac	NLQYAA	YP2A
21	6	Ac	NLQYDA	YP2D
22	6	-	NLQYFA	YP2F
23	6	Ac	NLQYFA	YP2F
24	6	Ac	NLQYIA	YP2I
25	6	Ac	NLQYLA	YP2L
26	6	Ac	NLQYRA	YP2R
27	6	Ac	NLQYVA	YP2V
28	6	Ac	NLQYWA	YP2W
29	6	Ac	NLQAYA	YP3A
30	6	Ac	NLQCYA	YP3C
31	6	Ac	NLQDYA	YP3D
32	6	-	NLQFYA	YP3F
33	6	Ac	NLQFYA	YP3F
34	6	Ac	NLQIYA	YP3I
35	6	Ac	NLQLYA	YP3L
36	6	Ac	NLQPYA	YP3P
37	6	Ac	NLQRYA	YP3R
38	6	Ac	NLQSYA	YP3S
39	6	Ac	NLQVYA	YP3V
40	6	-	NLQWYA	YP3W
41	6	Ac	NLQWYA	YP3W
42	6	-	NLGYYA	QP4G
43	6	Ac	NLGYYA	QP4G
44	6	-	NLSYYA	QP4S
45	6	Ac	NLSYYA	QP4S
46	6	Ac	NLAYYT	AP1T, QP4A
47	6	Ac	NLCYYT	AP1T, QP4C
48	6	Ac	NLDYYT	AP1T, QP4D
49	6	Ac	NLEYYT	AP1T, QP4E
50	6	Ac	NLFYYT	AP1T, QP4F
51	6	-	NLGYYT	AP1T, QP4G
52	6	Ac	NLGYYT	AP1T, QP4G
53	6	Ac	NLHYYT	AP1T, QP4H
54	6	Ac	NLIYYT	AP1T, QP4I
55	6	Ac	NLKYYT	AP1T, QP4K
56	6	Ac	NLLYYT	AP1T, QP4L
57	6	Ac	NLMYYT	AP1T, QP4M
58	6	Ac	NLNYYT	AP1T, QP4N
59	6	Ac	NLPYYT	AP1T, QP4P
60	6	Ac	NLRYYT	AP1T, QP4R
61	6	-	NLSYYT	AP1T, QP4S
62	6	Ac	NLSYYT	AP1T, QP4S
63	6	Ac	NLTYYT	AP1T, QP4T
64	6	Ac	NLVYYT	AP1T, QP4V

ID	Peptide length	N-term Modification	Sequence	Modification
65	6	Ac	NLWYYT	AP1T, QP4W
66	6	Ac	NLYYYT	AP1T, QP4Y
67	6	-	RLGYYA	QP4G, NP6R
68	6	Ac	RLGYYA	QP4G, NP6R
69	6	-	DLQYYA	QP6D
70	6	Ac	DLQYYA	QP6D
71	6	-	NPQYYA	LP5P
72	6	Ac	NPQYYA	LP5P
73	6	-	NFGYYA	QP4G, LP5F
74	6	-	NIGYYA	QP4G, LP5I
75	6	-	NRGYYA	QP4G, LP5R
76	6	-	NKGYYA	QP4G, LP5K
77	6	-	NDGYYA	QP4G, LP5D
78	6	Ac	RLSYYT	AP1T, QP4S, NP6R
79	6	Ac	NQSYYT	AP1T, QP4S, LP5Q
80	6	Ac	NRSYYT	AP1T, QP4S, LP5R
81	6	Ac	NKSYYT	AP1T, QP4S, LP5K
82	6	Ac	NESYYT	AP1T, QP4S, LP5E
83	6	-	NLSFYA	YP3F, QP4S
84	6	Ac	NLSFYA	YP3F, QP4S
85	6	Ac	NLSYFA	YP2F, QP4S
86	6	-	NLGYFA	YP2F, QP4G
87	6	Ac	NLGYFA	YP2F, QP4G
88	6	-	NLGFYA	YP3F, QP4G
89	6	Ac	NLGFYA	YP3F, QP4G
90	6	-	NLGHYA	YP3H, QP4G
91	6	Ac	NLGHYA	YP3H, QP4G
92	6	-	NLSHYA	YP3H, QP4S
93	6	Ac	NLSHYA	YP3H, QP4S
94	6	Ac	NLGYFT	AP1T, YP2F, QP4G
95	6	-	NLSFYT	AP1T, YP3F, QP4S
96	6	Ac	NLSFYT	AP1T, YP3F, QP4S
97	6	-	NLSYFT	AP1T, YP2F, QP4S
98	6	Ac	NLSYFT	AP1T, YP2F, QP4S
99	6	-	NLSFFT	AP1T, YP2F, YP3F, QP4S
100	6	Ac	NLSFFT	AP1T, YP2F, YP3F, QP4S
101	6	-	NLSY4T	AP1T, YP2 4, QP4S
102	6	-	NLS4YT	AP1T, YP3 4, QP4S
103	6	-	NLSY2T	AP1T, YP2 2, QP4S
104	6	-	NLSY3T	AP1T, YP2 3, QP4S
105	6	-	NLSY5T	AP1T, YP2 5, QP4S
106	6	Ac	NLSYAT	AP1T, YP2A, QP4S
107	6	Ac	NLSAYT	AP1T, YP3A, QP4S
108	6	-	NLQYHA	YP2H
109	6	-	NLAYYA	QP4A
110	6	-	NLEYYA	QP4E
111	6	-	NLLYYA	QP4L
112	6	-	NLNYYA	QP4N
113	6	-	NLPYYA	QP4P
114	6	-	NLRYYA	QP4R
115	6	-	NLTYYA	QP4T
116	6	Ac	NLSYHT	AP1T, YP2H, QP4S
117	6	Ac	NLSYWT	AP1T, YP2W, QP4S

Table S2.2. Cryo-EM data collection, refinement, and validation statistics.

PDB ID	6XMJ
EMDB ID	22259

Data collection and Processing	
Microscope	Titan Krios
Camera	K2 Summit
Magnification	29,000
Voltage (kV)	300 keV
Total electron fluence ($e^-/Å^2$)	50
Electron flux (e^-/pixel/sec)	8
Defocus range (μm)	-1.5 to -3.0
Pixel size (Å)	1.03
Micrographs collected (no.)	13,392
Total extracted particles (no.)	2,650,143
Refined particles (no.)	497,630

Reconstruction	
Final particles (no.)	234,960
Symmetry	C1
Resolution (global, Å)	
FSC 0.5 (unmasked/masked)	3.1/3.0
FSC 0.143 (unmasked/masked)	3.0/2.9
Resolution Range (local, Å)	2.8-3.6

Model composition	
Nonhydrogen atoms	64,832
Protein residues	4,555
Ligands	0
Waters	0

Refinement	
MapCC (volume/masked)	0.80/0.84
Map sharpening B factor ($Å^2$)	-108
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.746

Validation	
EMRinger score	3.97
CaBLAM outliers (%)	1.79
MolProbity score	1.13
Clashscore	1.77
Rotamer outliers (%)	0
Ramachandran plot	
Outliers (%)	0.02
Allowed (%)	3.19
Favored (%)	96.79

Table S2.3. Measurements associated with open-gate *h*20S cryo-EM structure.

	α-subunit						
	α1	α2	α3	α4	α5	α6	α7
Pro17 displacement (Å)	1.0	1.3	2.8	3.7	3.5	1.6	1.5
Rigid body rotation (°)	2.2	1.3	1.0	1.7	1.1	1.6	1.3
Unresolved gating residues (#)	8	7	7	2	7	3	6

	C-tail (α-pocket)						
	α7/α1	α1/α2	α2/α3	α3/α4	α4/α5	α5/α6	α6/α7
C-tail electron density (Å^3)	4.7	165.8	196.8	161.8	189.0	170.2	73.9
Salt bridging distance (Å)	-	5.1	3.4	3.2	2.9	3.6	4.3
P2-P3 ϕ (°)	-	-61.5	-127.3	-121.1	-137.8	-131.8	-129.9
P2-P3 ψ (°)	-	-39.8	159.2	167.2	149.2	136.3	150.9
π-stacking R (Å)	-	4.3	4.6	4.6	4.7	4.4	4.7
π-stacking θ (°)	-	25.7	26.0	26.3	23.7	33.3	25.6

	α-ring pore						
	α1-α4	α1-α5	α2-α5	α2-α6	α3-α6	α3-α7	α4-α7
Diameter (closed-gate) (Å)[a]	37.5	39.2	37.7	36.9	36.4	36.8	36.9
Diameter (open-gate) (Å)	41.1	41.0	40.9	41.1	41.2	41.2	41.2

[a]Diameters are defined as the average distance between αPro17 Cα atoms across the α-ring denoted by the α-subunit pairs in the *h*20S (PDB ID: 4R3O)[32].

Table S2.4. C-terminal sequences of PA200, PAN, Rpt5, Rpt3 and Rpt2 from different organisms identified in UniProtKB.

Protein	Entry	Sequence	Entry name
PAN	Q58576	LDVLYR	PAN_METJA
PAN	D4GUJ7	VSRAFA	PAN1_HALVD
PAN	Q5UT56	FTDYQY	PAN2_HALVD
PAN	Q9V2V6	DEQTEE	PSA1_HALVD
PAN	D4GYZ1	NFEGLE	PSB_HALVD
PAN	P9WQN5	NLGQYL	ARC_MYCTU
PAN	Q9HRW6	VSRTFA	PAN1_HALSA
PAN	Q9V287	HEIIYG	PAN_PYRAB
PAN	O50202	NTGQYL	ARC_RHOER
PAN	Q8TI88	ETTMFV	PAN_METAC
PAN	Q8PY58	PETMFV	PAN_METMA
PAN	Q9HNP9	YPSYIQ	PAN2_HALSA
PAN	Q9V2V5	TDEREE	PSA2_HALVD
PAN	O28303	KGVMFV	PAN_ARCFU
PAN	Q8U4H3	HEVIYG	PAN_PYRFU
PAN	O57940	HEVIYG	PAN_PYRHO
PAN	Q5JHS5	HEVMYG	PAN_THEKO
PAN	Q980M1	RREKYS	PAN_SACS2
PAN	C5A6P8	HEVMYG	PAN_THEGJ
PAN	Q975U2	RTEKYV	PAN_SULTO
PAN	Q6LWR0	LTVMYG	PAN_METMP
PAN	A9A916	LTVMYG	PAN_METM6
PAN	Q2FQ56	AGVMFA	PAN_METHJ
PAN	A6UQT3	LTTMYG	PAN_METVS
PAN	O26824	TGVMFG	PAN_METTH
PAN	A6VHR1	LTVMYG	PAN_METM7
PAN	B8GGN4	SGAMFA	PAN_METPE
PAN	C3MY47	RREKYS	PAN_SULIM
PAN	Q0W257	SGVMFA	PAN_METAR
PAN	A7I8B8	EGRMFA	PAN_METB6
PAN	Q8TX03	FKRAYH	PAN_METKA
PAN	C3N7K8	RREKYS	PAN_SULIY
PAN	Q9YAC7	TIATVI	PAN_AERPE
PAN	A4G0S4	LTVMYG	PAN_METM5
PAN	A3CV35	FGEMFA	PAN_METMJ
PAN	C3MRF1	RREKYS	PAN_SULIL
PAN	C3NFW6	RREKYS	PAN_SULIN
PAN	C4KIR6	RREKYS	PAN_SULIK
PAN	B6YXR2	HEVMYG	PAN_THEON
PAN	C3MZI6	RREKYS	PAN_SULIA
PAN	P9WQN4	NLGQYL	ARC_MYCTO
PA200	Q14997	SPCYYA	PSME4_HUMAN
PA200	Q5SSW2	SPCYYA	PSME4_MOUSE
PA200	F1MKX4	SPCYYA	PSME4_BOVIN
PA200	Q6NRP2	SPCYYA	PSME4_XENLA
PA200	F1QFR9	SPCYYA	PSM4A_DANRE
PA200	P43583	WRSYYA	BLM10_YEAST
PA200	F4JC97	SSSYFA	PSME4_ARATH
Rpt5	Q9SEI2	SLNYYA	PS6AA_ARATH
Rpt5	O76371	SLNYYA	PRS6A_CAEEL
Rpt5	Q54PN7	TLEYYA	PRS6A_DICDI
Rpt5	Q8SR13	KLLYFT	PRS6A_ENCCU
Rpt5	P17980	NLQYYA	PRS6A_HUMAN
Rpt5	O88685	NLQYYA	PRS6A_MOUSE
Rpt5	Q63569	NLQYYA	PRS6A_RAT
Rpt5	P33297	SVSFYA	PRS6A_YEAST
Rpt5	O42587	NLQYYA	PR6AA_XENLA

Protein	Entry	Sequence	Entry name
Rpt3	P43686	EHEFYK	PRS6B_HUMAN
Rpt3	P54775	EHEFYK	PRS6B_MOUSE
Rpt3	Q63570	EHEFYK	PRS6B_RAT
Rpt3	P33298	KFDFYK	PRS6B_YEAST
Rpt3	Q9SEI4	DFEFYK	PRS6B_ARATH
Rpt3	Q3T030	EHEFYK	PRS6B_BOVIN
Rpt3	O74894	QFAFYK	PRS6B_SCHPO
Rpt3	Q4R7L3	EHEFYK	PRS6B_MACFA
Rpt3	P46502	DFEFYK	PRS6B_CAEEL
Rpt2	P40327	LEGLYL	PRS4_YEAST
Rpt2	Q9SZD4	PEGLYM	PRS4A_ARATH
Rpt2	P62191	PEGLYL	PRS4_HUMAN
Rpt2	P62192	PEGLYL	PRS4_MOUSE
Rpt2	P48601	PEGLYL	PRS4_DROME
Rpt2	P62193	PEGLYL	PRS4_RAT
Rpt2	O16368	PEELYL	PRS4_CAEEL
Rpt2	Q90732	PEGLYL	PRS4_CHICK
Rpt2	Q8SRH0	SAGLYS	PRS4_ENCCU

Chapter 3

Enabling the discovery of novel regulators of the human 20S proteasome

Kwadwo A. Opoku-Nsiah, Sarah K. Williams and Jason E. Gestwicki

Abstract

The proteasome degradation system has evolved in eukaryotes to include a diversity of cell type-specific regulatory binding partners of the 20S proteasome (20S), enabling cells to spatially and temporally regulate protein turnover. As the biology of proteasomal degradation becomes further implicated in human disease, the need for small molecule modulators that can probe the 20S becomes increasingly urgent. In this chapter, we detail the progress of ongoing efforts to deepen our understanding of the YΦ motif through additional peptide studies of the 20S proteasome. In addition, we have outlined experimental strategies that leverage the development of recently developed probes that will enable further characterization of the human 20S gating mechanism. Finally, leveraging insights from our studies, we identified 10 proteins in the reference human proteome, with accessible C-terminal YΦ motifs, as potential binding partners and putative regulators of the 20S proteasome.

Introduction

Protein turnover by the proteasome degradation system is a critical arm of cellular proteostasis that also directly influences a diversity of disparate biological processes within eukaryotes. In turn, cells rely on several mechanisms to regulate the proteolytic activity of the system's core 20S proteasome complex (20S). For instance, proteasome activators (PAs) enhance proteasomal degradation with spatial and temporal specificity through HbYX motifs located at the extreme C-terminus. Conserved protein-protein interactions (PPIs) between he HbYX motif and pockets along the exterior of the 20S facilitate gate opening of its α-ring and subsequent activation. In our recent study, we

identified novel intra- and intermolecular interactions, along with the contacts described in the HbYX model, that enable activation of the human 20S (h20S). These findings have reshaped the canonical mechanism of proteasome activation and have been summarized in an updated model defined through the nYΦn motif. Further characterization of this mechanism may enable future discovery of pharmacological 20S activators as potential therapies against proteostasis-associated diseases.

Elucidation of the HbYX motif, over a decade ago, has facilitated the discovery of previously unknown regulators of the 20S and, in turn, novel aspects of proteasome biology through recognition of the conserved C-terminal sequence. Such findings include, the proteasome's role in endoplasmic reticulum-associated degradation (ERAD) via engagement with the HbYX motif-containing ATPase, p97[1]. More recently, proteasomal degradation has been linked to neuronal development and synapses through two separately studied adaptor proteins, pi31 and GPM6[2–4]. Unlike the canonical PAs, many HbYX-containing regulators, including these three, do not enhance proteasome function but regulate the subcellular localization of the 20S (and its complexes). Interestingly, PA200 is the only known binding partner capable of forming stable complexes with the 20S through a single C-terminal motif[5], whereas all other reported regulators mediate these C-termini-dependent PPIs as multivalent oligomers. Furthermore, PA200 is the only regulator with a bona fide YΦ motif, to suggest that the YΦ motif potentially binds with tighter affinity and can therefore form stable complexes, without avidity, between the 20S and its regulators. If true, the eukaryotic proteome may contain unannotated proteins with C-terminal YΦ motifs that potentially regulate the 20S.

Follow-up studies on the YΦ motif.

The recently characterized nYΦn motif centers on the interactions between the aromatic P2 and P3 side chains and the residues within the α-pocket, which are reportedly necessary for stimulation by monovalent and, in certain cases, multivalent PAs (**Figure 2.5B, C** and **Figure 2.6B**). However, the interactions provided by the residues flanking this central pharmacophore (P1 & P4) and those mediated by residues that are further upstream, also impact activity (**Figure 2.1B, D, & E**), and thus, warrant further elucidation in the development of more potent peptide activators of the *h*20S. We generated a chemically diverse series of P5-substituted Opt5 peptides (Ac-NLSYYT) and measured their ability to stimulate the *h*20S. Residues with hydrophobic side chains were significantly preferred at the P5 position, with Leu as the most stimulatory residue among those tested, suggesting van der Waal interactions along the pocket. In general, peptides with aromatic side chains (Trp, Tyr and Phe) at the P5 position were moderately more potent but displayed diminished activity relative to the original Opt5 (Leu) sequence (**Figure S3.1**).

With a more exhaustive panel of substitutions, we revisited the contributions of the P1 position in Opt5. As previously shown (see **Figure S2.4D**), Opt5 peptides display complete dose-response curves, relative to Rpt5 (data not shown), allowing for comparisons to be drawn based on relative potencies. Based on the activities, the peptides were binned into one of three categories: full agonist, partial agonist, and inactive. Inactive peptides had P1 residues with either charged and/or conformationally encumbered (Asp, His, Phe and Pro) side chains, which are likely to electrostatically or sterically impede the adjacent terminal carboxylate in forming the requisite salt bridge

with αLys66. Similarly, peptides that are partial agonists are unable to adopt the optimal orientations to forge the necessary contacts, including Gly or Asn at the P1 position. Alternatively, C-terminal Asn residues are prone to chemical modifications, which can potentially explain the diminished activity observed[6]. The two additional partial agonists contained hydrophobic P1 residues (Leu & Val) and reported significantly better potencies (EC_{50} = 69.2 ± 1.1 and 27.2 ± 1.3 μM, respectively). Furthermore, a P1 Ala fully activated the h20S, suggesting that polar interactions of the P1 side chain are not essential for binding or robust stimulation. However, small, polar residues were preferred at P1 (Ser and Thr), with Ser displaying a two-fold improvement in potency (EC_{50} = 17.3 ± 1.2 μM) over Thr (**Figure 3.1A**).

Although P4-modifications have the greatest magnitude effect on peptide activity, it remains unclear how the interactions at this position influences gate opening. We have previously shown that Opt5 peptides with a P4-Gly or P4-Pro (Opt5^{SP4G} and Opt5^{SP4P}), residues commonly associated with disfavoring α-helices[7], robustly stimulated the h20S (**Figure 2.1E**). Additionally, the Opt5 sequence was shown to bind with β-strand structure in five of the six occupied α-pockets (**Figure 2.3B**), suggesting a bias for adopting the β-strand conformation upon binding the h20S. Therefore, we speculated that modifications increasing the β-strand propensity of the Opt5 peptide should improve the relative potency of the peptide. Performing a proline mutational scan of the amino acid residues N-terminal of the YΦ motif, we identified that a P6 Pro (Opt5^{NP6P}) was 1.5 x more potent than WT Opt5; however, Pro-substitutions in Opt5 routinely diminished the capacity of the peptide to stimulate the h20S (**Figure 3.1B**). Finally, we attempted to disrupt the activity of Opt5 with the helical-promoting unnatural α-methylAla (AIB) residue at the P4 position

but observed the opposite effect. Furthermore, the peptide with a P4-AIB stimulated the *h*20S to a greater extent than Opt5^{SP4G} and Opt5^{SP4P}, demonstrating again how changes in peptide activity do not seem to reflect the predicted preference for β-strand over α-helical secondary structure (**Figure 3.1C**).

In addition to conferring structure to downstream residues, Pro-substitutions displace local H-bonding contacts, which may explain the role of the P4 residue in gate opening. Replacing the P4 residue with an N-methylated Ala, effectively suppressed the activity of the Opt5 peptide by (xx). H-bond interactions along the amide backbone seem to be critical for peptide activity as noted by a series of analogs that either swap the configuration of the stereocenters at the α-carbon with the *D*-amino acid enantiomer (**Figure 3.1D**) or the entire network of peptide H-bonds with the peptoid backbone (**Figure 3.1E**). From these peptide studies, we concluded that the activity of Opt5 peptides is not heavily influenced by the peptide's ability to adopt secondary structure. Our data suggests that the intermolecular H-bonding of the peptide backbone represents a significant enthalpic contribution to α-pocket binding. If true, good luck identifying a non-peptide scaffold that can meet the same enthalpic contributions or can compensate the enthalpic penalty.

In an attempt to contextualize our findings, we can revisit the snapshots of the α-pocket-Opt5 PPI in **Figure 2.3A** of chapter 2. Although we reported a strong preference for hydrophobic side chains at the P5 position, the P5 Leu was unresolved in the structure to suggest that this portion of the C-tail remains dynamic upon bonding. The α-pocket becomes more solvent-exposed around the P5 position; therefore, upon binding the α-pocket, Opt5 may rely on sufficient van der Waal interactions to ward off water molecules

from displacing Opt5 through some inverse of the hydrophobic effect. Due to the resolution of our structure, the exact orientation of the P1 residue is uncertain, however, it is plausible that the Ser or Thr side chains would pick up polar contacts within the α-pocket. Interestingly, the hydrophobic residues at the P1 position (Leu and Val) may be governed under the same principles linked to the hydrophobic effect. Indeed, Opt5 forms several H-bond contacts between its amide backbone and atoms throughout the α-pocket. Moving forward, we must exploit novel chemistries to modulate aspects of the P2-P3 pharmacophores in order to drive down potency and achieve cellular efficacy.

C-terminome reveals potential binders of the 20S.

Surveying the reference human proteome, we identified more than 500 proteins with C-terminal HbYX motifs, which included sequences containing a penultimate (P2) Phe or Tyr (Φ) residue and a hydrophobic amino acid residue at the P3 position. Of these hits, only 39 proteins had C-terminal sequences that agreed with the residue preferences denoted for the YΦ motif (see **Figure 2.6** in chapter 2). Cluster analysis followed by manual curation resulted in a final count of 10 proteins as potential binding partners of the *h*20S (**Figure 3.2A**). To preliminarily determine if any of these hits were PAs, we synthesized N-terminally acetylated hexapeptides corresponding to the C-termini of these proteins and measured their ability to stimulate the *h*20S. Unsurprisingly, none of the peptides were particularly stimulatory relative to Rpt5, but only 3 were completely inactive and exhibited activity below DMSO-treated *h*20S (**Figure 3.2B**).

The proteins identified in the C-terminome analysis are associated with one of three biological processes: p-granule function, RNA binding & processing, or

ubiquitination. Indeed, there is precedence for the sequestration of the 20S into p granules during cellular stress[8]. Furthermore, such phase-separated granules are commonly enriched with RNA-binding proteins as well as intrinsically disordered proteins[9], detailing a subcellular local environment where the degradative capacity of the 20S could, alone, temporally regulate gene transcription. However, our focus was immediately drawn to one of two E3 ubiquitin ligases that were identified from the C-terminome analysis, Mdm2.

Polyubiquitin tags are typically used as a signal for protein substrates to be targeted and degraded by the proteasome through the ubiquitin proteasome system (UPS), directly implicating Mdm2's canonical function to the 20S. However, experimental evidence has demonstrated that cellular turnover of several Mdm2 substrates, including p21, p53 and retinoblastoma-associated protein (Rb1), can be mediated by Mdm2 in a ubiquitin-independent manner[10–12]. Moreover, most substrates of Mdm2 are intrinsically disordered, containing long stretches of disordered regions (**Figure 3.2C**)[13], and thus, capable of being degraded by the free, unbound core 20S. Reportedly, the ubiquitin-independent degradation of p21 and p53 can be mediated by direct PPIs between the nuclear activator PA28γ. However, to date, a ubiquitin-independent mechanism explaining the turnover of Rb1 has yet to be proposed. Herein, we postulate that direct PPIs between the 20S and the C-terminal YΦ motif of Mdm2 may facilitate turnover of Rb1 (**Figure 3.2D**). This secondary route to the proteasome via Mdm2, may also facilitate the degradation of p21 and p53, and ultimately establishes an alternative mechanism for E3 ligases and potentially other classes of proteins to regulate the turnover of the disordered proteome.

Tools, strategies, and future directions.

Structural determination of the activated h20S complex was a crucial piece in resolving the YΦ motif and remains a linchpin in our efforts to better understand proteasome activation. One of the greatest hurdles with researching the human 20S was cost. None of this work would have been possible without Amnon and the micrograms of 20S proteasome that he purified from the blood of Israeli citizens. The ability to pursue future studies in the archaeal proteasome and/or with proteasomes originating from a cellular source would hasten progress.

When the PA26^{E102A} constructs were tested against the archaeal 20S, we observed remarkably similar SAR to the activities reported for the human 20S (**Figure S3.2A, B**). Furthermore, the measured binding affinities were also on par between the two proteasome samples (**Figure S3.2C**). One interesting distinction, was that PA26^{FF} did not partially activate the archaeal 20S as was reported with the h20S. This suggests that the intramolecular pi-stack may not be present or at least a prominent contributor to binding in archaea. Since many of the constructs tightly bind the archaeal 20S, acquiring a structure of the archaeal 20S-PA26^{YY} complex would quickly determine if the archaeal system would accurately inform on the YΦ motif.

Last but not least, we can elucidate the mechanism computationally. Performing molecular mechanics and docking-based studies are cost-effective and high-throughput approaches to interrogate the YΦ motif and potentially identify novel binders. Ironically, this is where we started minus a few important distinctions: a relevant structure, a refined model, a lil' wisdom, a highly capable successor, and of course, a PhD.

Materials and Methods

Archaeal *T acid.* 20S proteasome was provided by R. Huang from the laboratory of L.E. Kay (University of Toronto).

Peptoid FF-598 was provided by F. Fan from the laboratory of J.K. Sello (University of California San Francisco).

Peptide Synthesis. See chapter 2 (page 53).

Protein purification of PA26^{E102A} was expressed and purified as described in chapter 2 (page 54).

Proteasome activity assay. See chapter 2 (page 54). For archaeal 20S, 10 nM of *T acid.* 20S was used to perform activity assays in the presence of LLVY-amc (10 μM).

Biotinylation of *T acid.* 20S. See chapter 2, biotinylation of *h*20S (page 56).

Binding kinetic analysis of *T acid.* 20S. See chapter 2 (page 56).

References

1.   Isakov, E. & Stanhill, A. Stalled proteasomes are directly relieved by P97 recruitment. *J. Biol. Chem.* **286**, 30274–30283 (2011).

2.   Liu, K. *et al.* PI31 Is an Adaptor Protein for Proteasome Transport in Axons and Required for Synaptic Development. *Dev. Cell* **50**, 509-524.e10 (2019).

3.   Ramachandran, K. V & Margolis, S. S. A mammalian nervous-system-specific plasma membrane proteasome complex that modulates neuronal function. *Nat. Struct. Mol. Biol.* **24**, 419–430 (2017).

4.   Ramachandran, K. V. *et al.* Activity-Dependent Degradation of the Nascentome by the Neuronal Membrane Proteasome. *Mol. Cell* **71**, 169-177.e6 (2018).

5.   Sadre-Bazzaz, K., Whitby, F. G., Robinson, H., Formosa, T. & Hill, C. P. Structure of a Blm10 Complex Reveals Common Mechanisms for Proteasome Binding and Gate Opening. *Mol. Cell* **37**, 728–735 (2010).

6.   Friedrich, M. G., Wang, Z., Schey, K. L. & Truscott, R. J. W. Mechanism of protein cleavage at asparagine leading to protein–protein cross-links. *Biochem. J.* **476**, 3817–3834 (2019).

7.   Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* **12**, 6–15 (2012).

8.   Ustrell, Ë., Hoffman, L., Pratt, G. & Rechsteiner, M. PA200 , a nuclear proteasome activator involved in DNA repair. *EMBO J.* **21**, 3516–3525 (2002).

9.   Myers, N. *et al.* The Disordered Landscape of the 20S Proteasome Substrates Reveals Tight Association with Phase Separated Granules. *Proteomics* **1800076**,

1–9 (2018).

10.   Masson, P., Lundin, D., Söderbom, F. & Young, P. Characterization of a REG/PA28 proteasome activator homolog in dictyostelium discoideum indicates that the ubiquitin- And ATP-independent REGγ proteasome Is an ancient nuclear protease. *Eukaryot. Cell* **8**, 844–851 (2009).

11.   Zhang, Z. & Zhang, R. Proteasome activator PA28γ regulates p53 by enhancing its MDM2-mediated degradation. *EMBO J.* **27**, 852–864 (2008).

12.   Sdek, P. *et al.* MDM2 Promotes Proteasome-Dependent Ubiquitin-Independent Degradation of Retinoblastoma Protein. *Mol. Cell* **20**, 699–708 (2005).

13.   Guo, Z., Wang, X., Li, H. & Gao, Y. Screening E3 Substrates Using a Live Phage Display Library. *PLoS One* **8**, 1–12 (2013).
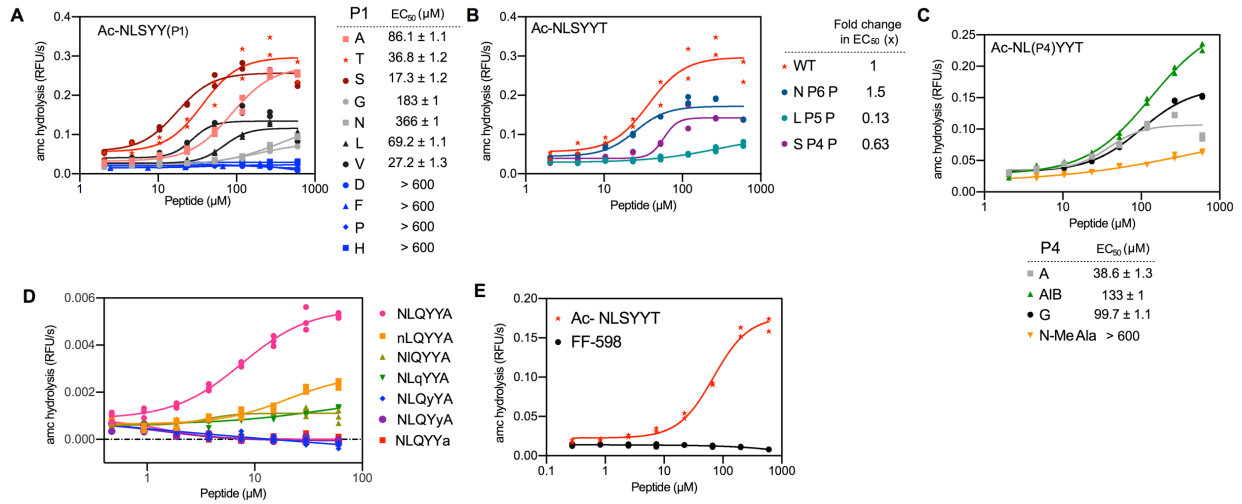
Figure 3.1. Native main and side chain H-bonds drive peptide activity.
A-E, Representative dose-response curves of P1-modified (**A**), Pro substitution scan (**B**), P4-modified with nonnatural amino acid residues (**C**), L to D enantiomers (**D**) and peptoid (**E**) with technical replicates plotted individually (n = 2 or 3). EC_{50} values are reported with error reported as s.e.m. (n = 2).
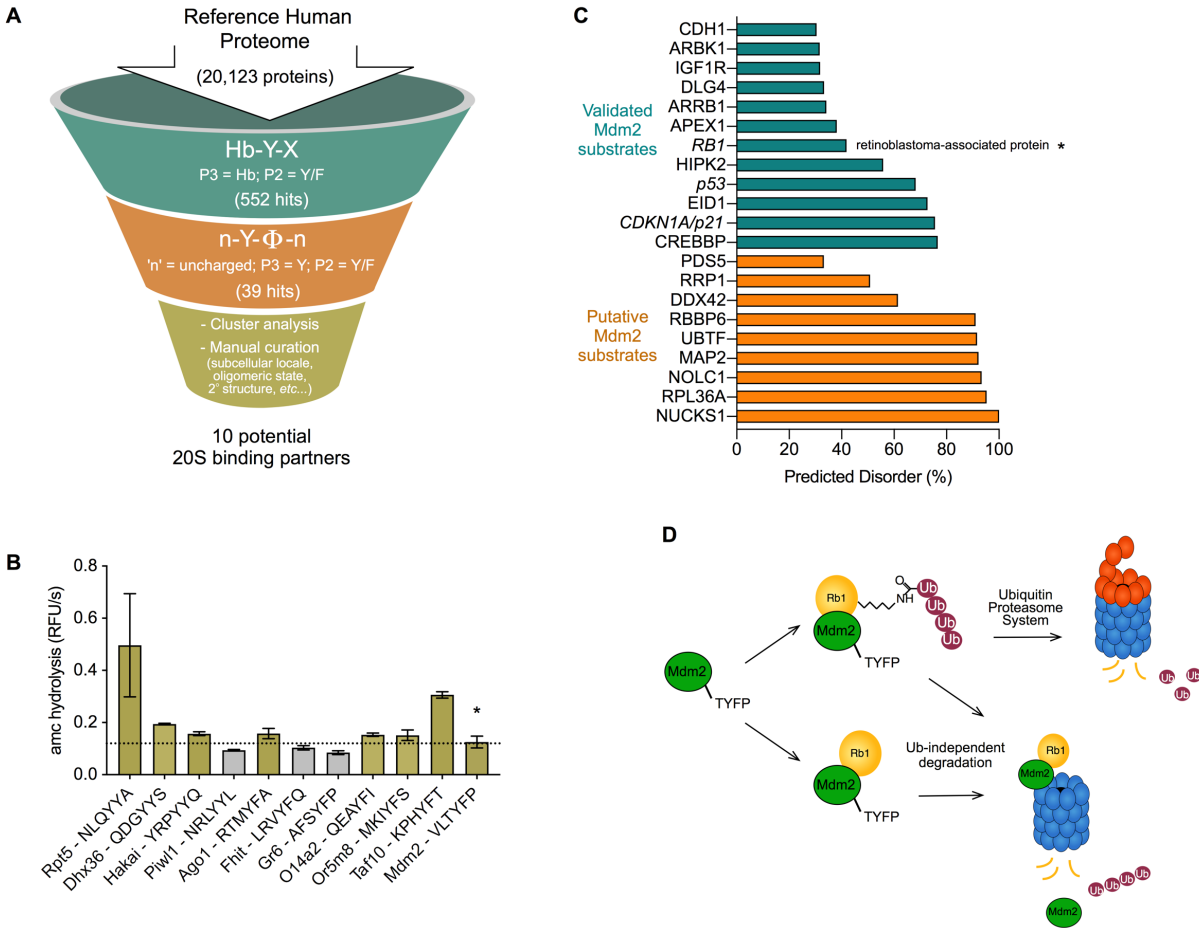
Figure 3.2. C-terminome studies reveal potential binding partners of the *h*20S.
A, Identification of ten proteins from the human proteome with C-terminal YΦ motifs that potentially target the *h*20S. See **Table S3.2** for detailed list. **B**, C-terminal YΦ hexapeptides (300 μM) of the hit proteins are modestly stimulators of the *h*20S *in vitro*. Representative rate of LLVY (10 μM) cleavage is reported for technical replicates (n = 3) with error reported as SD (n = 3). Active peptides (gold) stimulated above vehicle (dotted line). Mdm2 is denoted (*) **C**, Mdm2's validated (green) and predicted (orange)[13] substrates comprise of highly disordered proteins. Retinoblastoma-associated protein (Rb1) is denoted (*). Percent disorder was computed with www.pondr.com **D**, Proposed model for the ubiquitin-independent proteasomal degradation of Rb1 via Mdm2's YΦ motif.
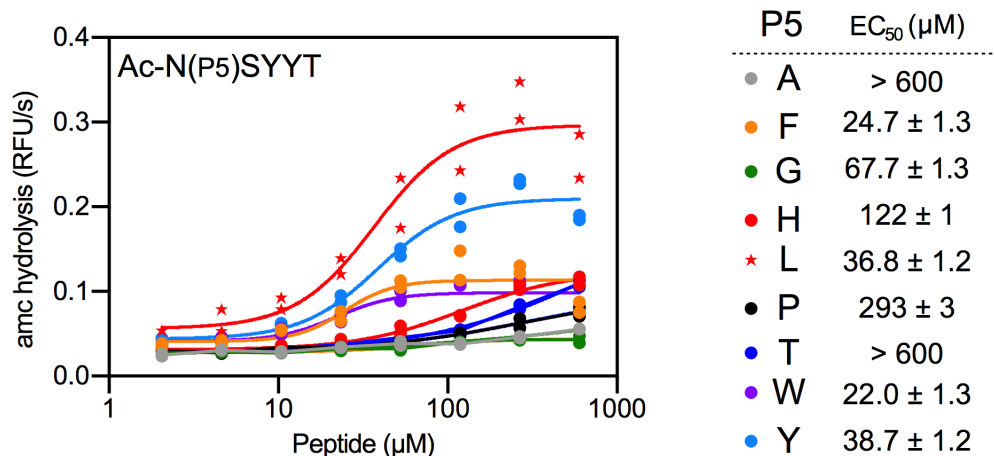
Figure S3.1. Hydrophobic side chains at P5 position enhance peptide activity. Stimulation of h20S by P5-substituted hexapeptides. Representative dose-response curves with technical replicate plotted individually (n = 2). EC_{50} values are reported for an independent experiment with error reported as s.e.m. (n = 2).
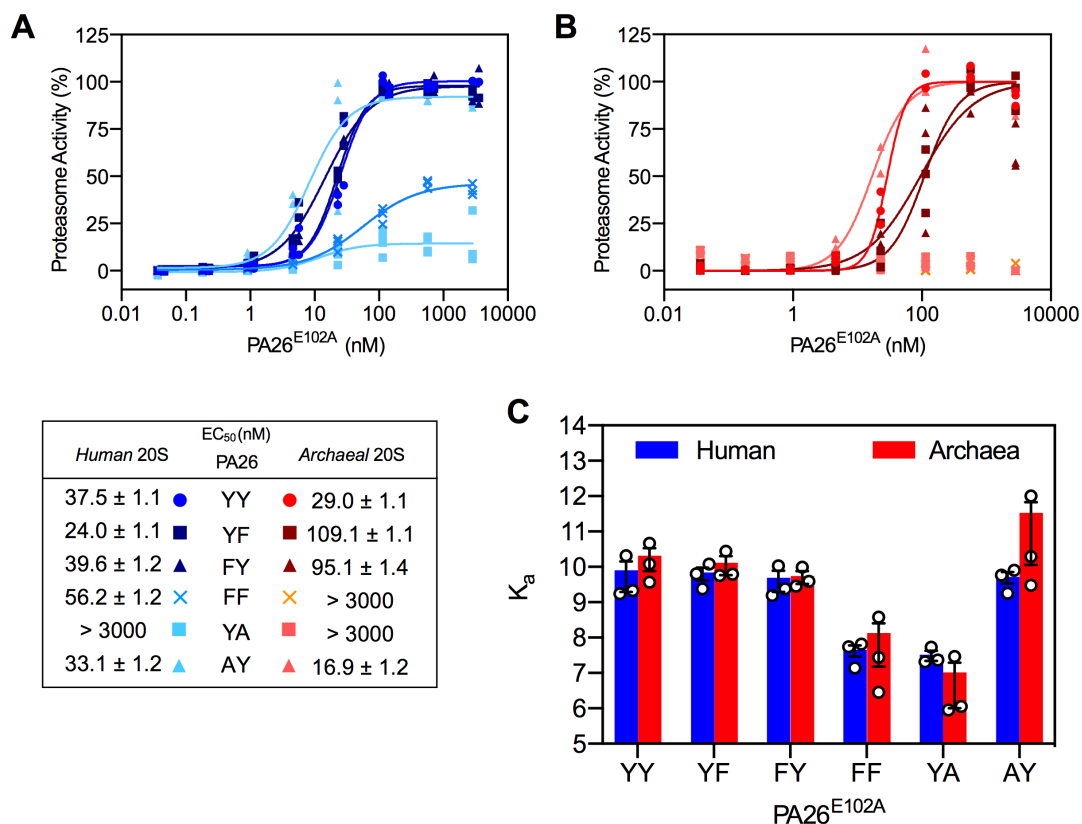
Figure S3.2. SAR of PA26's YΦ motif is consistent between human and archaea.
A, B, Proteasome activity of heptameric PA26[E102A-Opt5] constructs against the *h*20S (**A**) and archaeal 20S (**B**). **A, B**, Data are normalized PA26[YY] and plotted individually (n = 2 or 3). Reported EC_{50} is a mean of EC_{50} values calculated from two or three independent experiments with error reported as s.e.m. **C,** Reported association constants (K_a) for PA26[E102A] constructs binding to *h*20S (blue) or archaeal 20S (red). Binding kinetics were measured by BLI and mean K_a for experimental replicates (n = 3) are plotted with error reported as s.e.m. **B, C**, Sensorgram traces for *h*20S binding (see **Figure S2.11**) and raw activity data for *h*20S (see **Figure S2.4**) are reported in chapter 2.

Table S3.1. Manually curated list of YΦ proteins.[+]

Name	Descriptor	Localization	C-term Seq	C-term Exposed?	Oligomeric state	Size (aa)	Notes
DHX36	ATP-dependent RNA helicase DHX36 (DEAH box protein 36)	Nuc; Cyto; Chromosome; telomere; stress granules.	QDGYYS	NA	-	1008	Highly expressed in testis; Pro-rich C-term; interacts with Ago1 and 2
HAKAI	E3 ubiquitin-protein ligase Hakai (RING finger protein 188)	Nucleus speckle; Nuc; nucleoplasm	YRPYYQ	NA	homodimer	491	Pro-rich C-term; Baseline/differential expression (BDE)
PIWL1	Piwi-like protein 1	Cyto; P granule; chromatoid body;	NRLYYL	NA	NA	861	Endoribonuclease; Isoforms 3 & 4 in FY
AGO1	Protein argonaute-1 (Eukaryotic translation initiation factor 2C 1)	Cyto; P-body	RTMYFA	No	monomer	847	RNA-mediated gene silencing; disordered C-term (putative RNA-binding protein)
FHIT	Bis(5'-adenosyl)-triphosphatase (AP3A hydrolase) (AP3Aase)	Cyto; Mito; Nuc	LRVYFQ	Yes	homodimer	147	Phosphorylation of Tyr 145 by Src; low level of expression in all tissues; interacts with MDM2
GR6	Protein GR6	-	AFSYFP	NA	NA	149	Expressed in fetus 7 to 8 weeks
O14A2	Olfactory receptor 14A2 (Olfactory receptor OR1-31)	Cell mem; Multi-pass mem protein	QEAYFI	Cyto	NA	314	C-term is cytoplasmic;
OR5M8	Olfactory receptor 5M8 (Olfactory receptor OR11-194)	Cell mem; Multi-pass mem protein	MKIYFS	Cyto	NA	311	C-term is cytoplasmic;
TAF10	Transcription initiation factor TFIID subunit 10 (STAF28)	Nuc	KPHYFT	Yes	NA	218	Disordered C-term
MDM2	E3 ubiquitin-protein ligase Mdm2 (p53-binding protein Mdm2)	Nuc; nucleoplasm; Cyto; nucleolus	VLTYFP	Yes	Oligomer (MDM2, MDMX, FHIT)	491	Amplified expression in cancers; not in normal tissue; expressed ubiquitously; shown to interact w/ PSMA3

[+]Hits with inactive C-terminal peptides are denoted in grey.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Kwadwo Opoku-Nsiah

6592F07C233A4B6...   Author Signature

12/18/2020

Date