BMC Genomics

DATABASE

Open Access

ProEnd: a comprehensive database for identifying HbYX motif-containing proteins across the tree of life

David Salcedo-Tacuma[1], Giovanni D. Howells[1], Coleman McHose[1], Aimer Gutierrez-Diaz[2], Jane Schupp[1] and David M. Smith[1,3*]

Abstract

The proteasome plays a crucial role in cellular homeostasis by degrading misfolded, damaged, or unnecessary proteins. Understanding the regulatory mechanisms of proteasome activity is vital, particularly the interaction with activators containing the hydrophobic-tyrosine-any amino acid (HbYX) motif. Here, we present ProEnd, a comprehensive database designed to identify and catalog HbYX motif-containing proteins across the tree of life. Using a simple bioinformatics pipeline, we analyzed approximately 73 million proteins from 22,000 reference proteomes in the UniProt/SwissProt database. Our findings reveal the widespread presence of HbYX motifs in diverse organisms, highlighting their evolutionary conservation and functional significance. Notably, we observed an interesting prevalence of these motifs in viral proteomes, suggesting strategic interactions with the host proteasome. As validation two novel HbYX proteins found in this database were experimentally tested by pulldowns, confirming that they directly interact with the proteasome, with one of them directly activating it. ProEnd's extensive dataset and user-friendly interface enable researchers to explore the potential proteasomal regulator landscape, generating new hypotheses to advance proteasome biology. This resource is set to facilitate the discovery of novel therapeutic targets, enhancing our approach to treating diseases such as neurodegenerative disorders and cancer.

Keywords Proteasome, Proteostasis, HbYX, Database, Proend, Proteasome activation, Virus, Proteasome binding proteins, PIPs

Introduction

The proteasome is a key component in the cellular machinery, orchestrating the degradation of misfolded, damaged, or unnecessary proteins. Its function is critical not only for maintaining cellular homeostasis but also for regulating several biochemical pathways. In healthy cells, the proteasome ensures the proper turnover of proteins, preventing the accumulation of potentially toxic proteins that could disrupt cellular functions [1–3]. Conversely, in disease states, particularly in neurodegenerative diseases (ND) and certain cancers, the proteasome's role becomes even more crucial. Aberrations in proteasome activity

*Correspondence:
David M. Smith
dmsmith@hsc.wvu.edu
[1]Department of Biochemistry and Molecular Medicine, West Virginia University School of Medicine, 4 Medical Center Dr, Morgantown, WV, USA
[2]Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala 75007, Sweden
[3]Department of Neuroscience, Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV, USA

can lead to the accumulation of protein aggregates, a hallmark of many NDs, or affect the cell cycle and apoptosis, as observed in cancer [4–7].

Because of this pivotal role/importance in disease, in the last decade several groups have been trying to generate drugs/pharmaceuticals to specifically target the proteasome and modulate its function in hope to present a promising strategy for therapeutic intervention [7–9]. Velcade (Bortezomib), a pioneering proteasome inhibitor, exemplifies the proteasome's potential as a drug target. Approved for the treatment of multiple myeloma and lymphoma, Velcade capitalizes on the need of proteasome activity for cell survival, particularly in cancer cells with high rates of protein synthesis and turnover [10–12]. This success increased the interest in researching and developing proteasome activators or alternative inhibitors that could modulate proteasome activity, offering new avenues for treating NDs and cancer. The activation of the proteasome, for instance, could enhance the clearance of protein aggregates, a shared feature of NDs, potentially improving disease symptoms or progression [7, 8, 13, 14]. Thus, understanding the mechanisms of proteasome regulation, including the action of specific activators, inhibitors, and regulatory complexes, is vital for therapeutic intervention.

The 20S proteasome is defined as the core particle responsible for the proteolytic breakdown of proteins. Its cylindrical structure is composed of four stacked rings: two outer α-rings and two inner β-rings, each containing seven distinct subunits in the case of eukaryotes [15–18]. The catalytic activity resides within the β-rings, where three types of active sites, caspase-like, trypsin-like, and chymotrypsin-like facilitate the breakdown of proteins into peptides [15–17]. A critical aspect of the 20S proteasome's function is its gate, formed by the N-termini of the α-subunits. This gate controls access to the inner chamber, where protein degradation occurs, hence regulating substrate entry. The gate's default state is closed, preventing random protein degradation and ensuring selectivity in the proteasome's degradative capacity [15–17, 19].

Another system for regulation of substrate entry is the existence of proteasome activators, such as the 19S regulatory particle (PA700), PA200, and PA28γ (also known as 11S), which bind to the 20S core and induce conformational changes that open the gate [20–23]. The 19S regulatory particle, in particular, recognizes ubiquitinated proteins and uses ATPase activity to unfold and translocate them into the 20S core for degradation. Similarly, proteasome activators like PA200 or PA28γ enhance the proteasome's ability to degrade specific non-ubiquitinated substrates, highlighting the versatility and adaptability of proteasome regulation [20, 24].

These mechanisms to regulate proteasome activity exhibit remarkable conservation throughout evolution, spanning from archaea to eukaryotes. This evolutionary continuity allows us to observe specific features that are essential across different domains of life [25–27]. For instance, the archaea e.g. *Archegolbulous fuldgidous* contains a complete proteasome system including a 20S proteasome and an ATP-dependent proteasome activator known as PAN (proteasome-activating nucleotidase), which plays a pivotal role in this system, analogous to the function of the 19S regulatory particle in eukaryotic proteasomes [28]. Similarly, VCP (valosin-containing protein) also known as p97 or CDC48 in eukaryotes, has homologs in archaea (i.e. VAT) that shares functional similarities with PAN in its role in stimulating proteasome dependent protein degradation, highlighting the conserved nature of these proteolytic processes [29, 30].

The T20S proteasome (20S from *Thermoplasma acidophilum*) and its interaction with PAN have provided a simplified model to study the fundamental aspects of proteasome regulation. PAN, through its ATP binding activity, facilitates the opening of the 20S proteasome gate, enabling substrate entry, crucial for the degradation mechanism to function [31, 32]. This interaction between PAN and the 20S emphasizes the importance of ATP-driven conformational changes for proteasome activity, a concept that is universally observed across species. Moreover, VCP/p97 mediates the extraction of misfolded proteins from the ER for delivery to the proteasome for degradation, emphasizing the integrated nature of cellular degradation pathways [30, 33–35]. Through the examination of the basal archaeal proteasome system and its eukaryotic counterparts, we have gained invaluable insights into the conserved mechanisms of proteasome degradation. Moreover, this evolutionary perspective increases our comprehension of cellular regulation but also highlights potential for discovery of therapeutic targets that span across species that await discovery.

The elucidation of the regulatory mechanisms governing the interaction between the proteasome activators and the 20S proteasome has been significantly advanced by the discovery of the HbYX motif. This tripeptide motif, characterized by a hydrophobic amino acid (H), followed by a tyrosine (Y), and any amino acid (X), which must be the C-terminal residue, has emerged as a key facilitator of proteasome activation and function. Pioneering work on the archaeal PAN complex and subsequent studies elucidated and highlighted the motif's crucial role in gate opening, a necessary step for substrate entry into the 20S core [20]. The interaction of the HbYX motif with the 20S proteasome is mediated by its docking into specific pockets between adjacent α-subunits, known as α-pockets or intersubunit pockets, triggering allosteric changes that lead to gate opening, which have been recently characterized in mechanistic detail by Chuah et al. 2023 and Gestwiki et al. 2022 [36–38].

What makes the HbYX motif particularly intriguing is its prevalence among most proteins that bind to the proteasome [20]. This motif is not only conserved across a wide range of proteasomal ATPases, including those integral to the 19S regulatory particle, but also found in proteins such as PA200, PI-31, Pac-1, Pac-2, and p97. This conservation from archaea to humans highlights the motif's fundamental role in proteasome binding and proteasome activation (gate-opening) denoting both, its utility and specificity in regulating protein degradation.

The question arises as to whether we can identify other HbYX proteins to discover new pathways for protein degradation. In that context the use of bioinformatics has revolutionized our ability to uncover and characterize novel molecular entities, significantly advancing our understanding of complex biological systems. Protein domains and transcription factor motifs, crucial for regulating gene expression and protein function, serve as prime examples of targets identified through computational approaches [39, 40]. However, the journey toward discovering specific proteasome regulators, and specifically those associated with the HbYX motif, presents unique challenges within the bioinformatics landscape.

Standard bioinformatics tools and pipelines, including the widely used BLAST or MEME (Multiple Em for Motif Elicitation) suite, offer powerful means to search for and analyze motifs within protein sequences [41, 42]. These classical tools have been instrumental in identifying recurring patterns that play significant roles in protein interactions, localization, and function. Despite their utility, these tools often encounter limitations when tasked with detecting highly specific motifs like the HbYX. The inefficiency largely stems from the sheer diversity and complexity of protein length sequences. In addition, researchers often face challenges related to the structure of the data and the interpretation of outputs from these bioinformatics tools, requiring extensive manual curation to confirm the identified motifs.

In response to these challenges, there is a growing need for the development of specialized pipelines and tools [43] tailored to the detection and analysis of specific proteasome regulators. Such advancements would enhance the efficiency and accuracy of motif discovery processes, enabling researchers to more effectively sift through the vast universe of proteins for potential therapeutic targets. Despite the hurdles, the potential to uncover novel regulators of the proteasome through bioinformatics and novel machine/deep learning methods remains a promising frontier in molecular biology and disease research.

In this report we introduce the **Pro**teasome **En**d regulators **d**atabase- ProEnd, a comprehensive database designed to navigate the universe of proteins for the discovery of new molecules featuring the HbYX motif. ProEnd, represents both a curated source and specific tool to uncover and characterize the landscape of proteasomal regulators. This ProEnd tool is specifically engineered to address the limitations encountered with standard motif search methodologies. By optimizing the search to hone in on the unique characteristics of the C-terminal HbYX sequence, ProEnd has successfully identified a set of HbYX proteins across a wide range of organisms, from archaea to humans, constituting a specialized database dedicated to proteins with HbYX motif. This database serves as a foundational resource for researchers worldwide, facilitating searching, hypothesis generation, and exploration of the HbYX motif's role in proteasome regulation and cellular function. By providing a platform for systematic analysis and cross-species comparison, we enable the scientific community to advance the understanding of new regulators and the diverse functions of the HbYX motif.

Methods

Enrichment analysis

Enrichment analysis was performed for jawed vertebrate species by extracting lists of HbYX proteins and analyzing them in R (version 4.1.2) with the 'clusterProfiler' package [44]. The 'enrichGO' function was employed for each species, and a heatmap was created using -log10 transformed data to highlight significant enrichments in molecular functions.

MSA and HMM logo construction

Sequences of HbYX proteins were identified with ProEnd, retrieved from the UniProt database and aligned using Jalview [45] with the ClustalX alignment tool. For the construction of Hidden Markov Model (HMM) profiles, we utilized WebLogo3 [46] and Python library LogoMaker to generate sequence logos and visualize the conservation and variability of the HbYX amino acids. For PAN-2 (MJ1494), the substitution matrix BLOSUM62 was calculated using Python Bio. Align to assess evolutionary divergences and similarities with PAN.

HbYX containing protein plasmid construction and proteasome purifications

The DNA sequence for the HbYX protein PAN-2 (MJ1494) was fetched from UniProt and cloned into pET-28a(+) vectors for recombinant *E. coli* overexpression under a T7 promotor and lac operator. A 6His tag was added to the N-terminus of each construct for Ni-NTA purification and pulldown experiments. The 6His tag is followed by a TEV protease sequence (ENLYFQ/M) where "/" indicates the cut site. The pET28a(+) vector was chosen due to stable expression of other archaeal proteins in the past in our lab. DNA sequences were codon optimized for expression in the bacterial strain BL21 star DE3. Transformed bacteria are overexpressed

through IPTG induction. *T. acidophilum* 20S proteasomes were expressed using the pET-28a(+) vector and contain a C-terminal 6His tag on the β subunits preceded by a TEV protease sequence in order to remove the 6His tag following purification. The T20S was purified following Seemüller et al. [47], with variations. Following purification, the T20S was TEV cut. Following TEV cleavage the T20S was added to an Ni-NTA column and only the flow through was collected ensuring all T20S used in pulldown experiments are tagless.

Proteasome activity assays–peptide substrates

Fluorogenic substrate peptide, suc-LLVY-amc, was obtained from BostonBiochem. For activity assays, this substrate was dissolved in DMSO and used at a final concentration of 100 μM. The reaction mixture also included a final DMSO concentration of 2%. The assays were performed in a buffer at 45 °C, with 2 mM ATP and 10 mM MgCl2, to a total volume of 0.1 ml of reaction buffer. The assays were conducted over 30 to 60 min and the results were analyzed using BioTek Gen5 Data Analysis software. Fluorescence at excitation/emission 380/460 nm was measured every 55 s for 2 h. As LLVY bypasses the T20S gate the initial rate of increase in fluorescence intensity is directly proportional to proteasome quantity and activity.

Cell culture

HEK-293 cells were grown in DMEM (Life Technologies, Carlsbad, CA) supplemented with 10% FBS (Sigma, St. Louis, MO), and 1% penicillin/streptomycin (Life Technologies). Cells were cultured at 37 °C in 5% CO2 environment at a density of 5×106 cells/ml and collected to for pulldown assays. All cells were obtained from ATCC (Manassas, VA).

Ni-NTA pulldown assay

Cell lysates were clarified by centrifugation at 20,000 rpm using a F21-8×50y rotor (ThermoFisher) for 30 min. Next, 200μL of Ni-NTA agarose resin (Qiagen) is added to the supernatant and allowed to bind for 1 h shaking at room temperature. The resin is washed extensively with Wash buffer (50mM Tris+100mM NaCl supplemented with 2mM ATP, 10mM $MgCl_2$. Following this washing step the columns are capped and concentration matched T20S is added and allowed to incubate for 20 min at room temperature. After 20 min the column is uncapped and the flow-through of unbound proteins is allowed to flow through the column. Columns are then washed with buffer (1x column bed volume) to further remove unbound proteins. Finally, HbYX containing proteins along with HbYX bound proteins are eluted using the Wash buffer supplemented with 150mM imidazole. Presence of 20S in the elution fractions indicates binding to the HbYX protein bound on the Ni-NTA column, and the amount of 20S is quantified in the elution fraction by running a proteasome activity assay using the fluorogenic substrate suc-LLVY-amc as well as SDS-PAGE and western blot in the case of neurotensin.

SDS-PAGE and western blot

Proteins were separated by SDS-PAGE using Invitrogen NuPAGE 4–12% Bis-Tris protein gels following manufacturer's protocol. Samples were mixed with Invitrogen SDS sample buffer (4x). Protein bands were visualized with Coomassie stain (Simply Blue Safe Stain, Novex) according to manufacturer instructions. Protein band sizes (kDa) were visualized using the SeeBlue Plus2 prestained protein standard (ThermoFisher). For Western blot, the gel was transferred to an Immobilon®-FL PVDF membrane at 30V overnight. The membrane was blocked in Tris-buffered saline with 0.01% Tween-20 (TBST) and 10% nonfat milk for 1 h at room temperature (~20˚C), then briefly washed and incubated with the primary antibody (Anti-Proteasome α's1,2,3,5,6,7 subunits, Enzo, MCP231; Anti-PSMD7/Mov34, ab140428) diluted 1:1000 in TBST with 5% nonfat milk at 4˚C overnight. After washing three times for 5 min each with TBST, the membrane was incubated with the secondary antibody (DyLightTM 550, Thermo, 10173; Alexa Fluor Plus 680, A32729) diluted 1:3000 in TBST with 5% nonfat milk for 1 h at room temperature. The membrane was washed again three times for 5 min each with TBST and then imaged using an Amersham Typhoon (GE).

Alphafold multimer predictions

Structural models of PAN-2 (MJ1494; UniProt ID: Q58889) were generated using the AlphaFold Multimer [48] v2.3.295, specifically configured to predict multimeric assemblies, reflecting the protein's anticipated homohexameric formation. This computational modeling was performed on a high-performance computing platform, utilizing six chains to represent PAN-2oligomeric state accurately. The analysis employed a reduced database preset, with a template search cutoff incorporating the most current structural data available up to August 1, 2023, ensuring the model's relevance with the latest updates. The resulting models were visualized and refined using ChimeraX software.

Results

ProEnd in the UniProt/Swissprot reference universe

To address the question of the prevalence of HbYX motif-containing molecules and to overcome the limitations of existing bioinformatics tools, we developed a workflow designed to efficiently identify proteins with the specific C-terminal HbYX motif. Utilizing AWK, a text-processing tool, our pipeline operated through a three-step process that leverages protein data repositories for

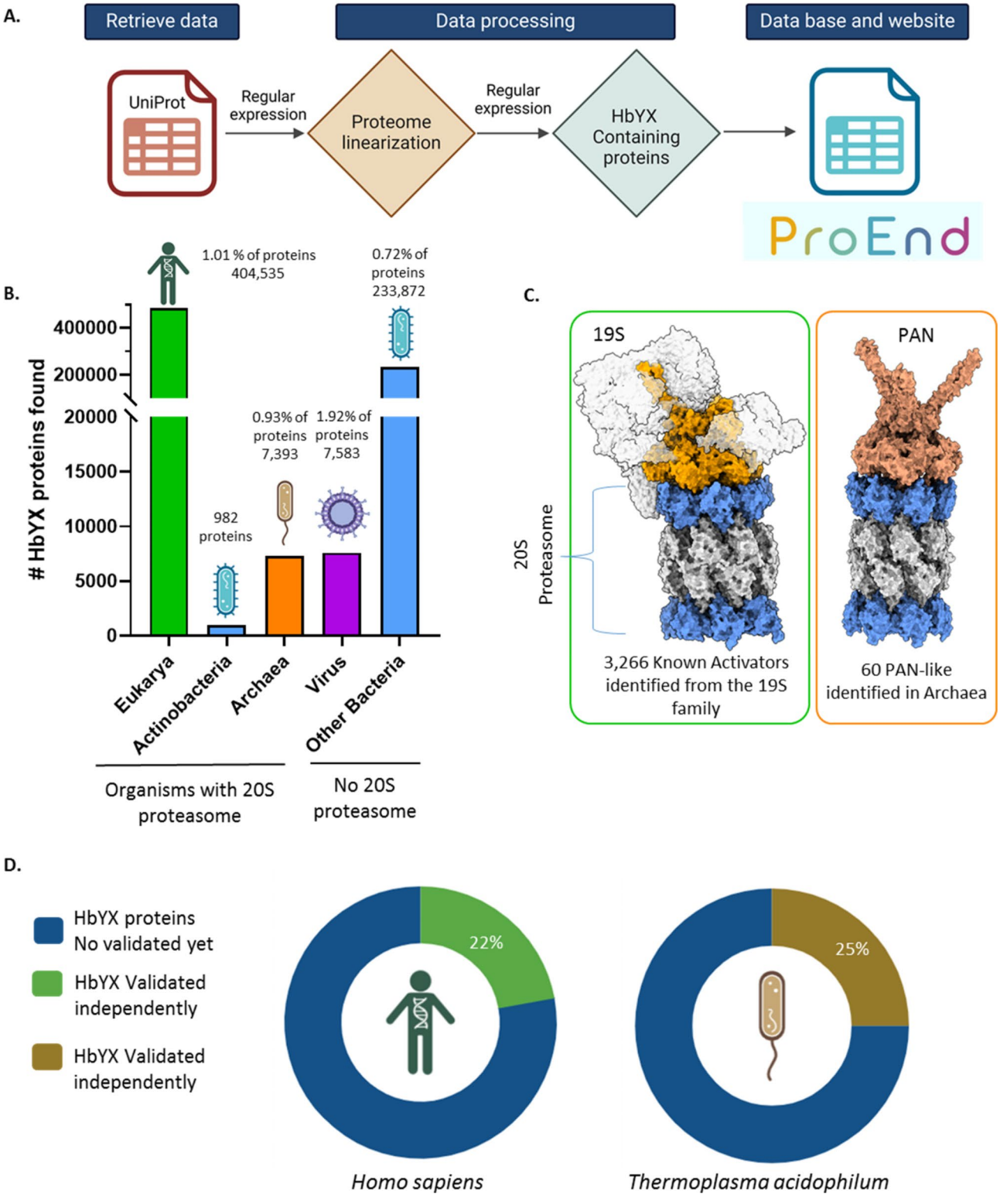

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Prevalence of HbYX motif across organisms identified by ProEnd. **A.** Workflow for the design of the ProEnd database and the pipeline used to retrieve proteomes from UniProt/SwissProt. AWK was utilized to develop the regular expressions. **B.** Distribution of HbYX Proteins Across Life Domains and Viruses. Barplot indicates the number of HbYX proteins within various life domains, represented as percentages indicating the average occurrence of HbYX proteins per domain. On the left side, organisms possessing a 20S proteasome—specifically Eukaryotes, Actinobacteria, and Archaea—are grouped as highlighted by the line below. On the right, organisms lacking a 20S proteasome, including viruses and other bacteria, are displayed. **C.** Number of known proteasome activators found in the ProEnd database. Structural representation of a 20S proteasome with α-rings in blue and β-rings in gray. The green box illustrates the eukaryotic 19S regulatory particle complexed with the 20 S proteasome (PDB: 6MSB), highlighting the conserved ATPases in orange. Additional regulatory proteins are outlined to enable comparison with the archaeal proteasome regulator. The orange box displays the structural complex of the archaeal 20S proteasome and PAN (PDB: 6HEC), with the PAN cap shown in salmon. In total 3,266 known activators form the 19S were identified in Eukarya and 60 PAN-like proteins were identified in Archaea. **D.** Percentage of HbYX proteins independently validated as proteasome interactors in other studies for *Homo sapiens* and *Thermoplasma acidophilum*, an archaeal model organism.

comprehensive analysis. First, we retrieved all reference proteomes from the UniProt [49], representing all available spectrum of organisms. Subsequently, we streamline the proteome data into a simplified format, arranging it into a single file where each protein is listed on a separate line (Linearization). In the final step of our pipeline, we deploy a carefully designed regular expression to sift through the linearized array of proteomes. This search was finely tuned to detect the presence of the HbYX motif, which is only found on the C-termini of the protein with the C-terminal residue being the "X" residue, ensuring that any occurrences are accurately captured. Upon identification, the protein and its associated motif sequence were cataloged in a structured table. Additionally, we developed a user-friendly online database (Fig. 1A), which not only enables researchers to interact with the dataset but also allows direct access to structure prediction data from Google DeepMind's AlphaFold and to UniProt entries by clicking on the protein ID. This approach not only simplifies the identification process but also facilitates a more manageable analysis of the data, allowing for a clearer understanding of the distribution and diversity of HbYX motif-containing proteins across different species.

Analysis of HbYX proteins

Our comprehensive analysis covered approximately 73 million proteins across about 22,000 reference proteomes within the UniProt/Swissprot databases. The goal was to catalog all proteins featuring the HbYX motif. From this extensive survey, we observed an interesting and noteworthy phenomenon in viruses. While more than half of the viral proteomes examined do not possess any proteins with the HbYX motif, contrarily the other half of viral proteomes contain a high prevalence of HbYX motif-containing proteins compared to other organisms. Specifically, we identified approximately ~7,500 HbYX proteins in viruses, which represent 1.92% of all proteins in these entities (Fig. 1B). Since a major function of the proteasome is to generate viral peptides for detection by the immune system (MHC-class I) [50, 51], these observations may indicate that virus could have evolved to either strategically avoid the proteasome or to negatively regulate it [52].

In addition, bacteria, which do not have a proteasome system, presented approximately 235,000 HbYX motif-containing proteins, accounting for 0.72% of the proteins analyzed in these organisms. In bacteria, protein degradation is primarily performed by ClpP or HslV proteases and auxiliary AAA-ATPase particles such as ClpA, ClpX, and HslU. These complexes are compartmentalized proteases like the eukaryotic proteasome but are not evolutionarily related [53, 54]. Although some studies suggest HslU/V is somewhat related to the proteasome system, and the C-terminus of HslU does bind to HslV in an ATP-dependent manner for regulatory purposes, which is analogous to 19S–20S or PAN-20S association [55, 56]. However, HslU does not have a HbYX motif [55, 56].

In addition, as a group, these bacterial protease systems typically rely on adaptor molecules to target proteins for degradation [57, 58]. The presence of HbYX motifs within bacterial proteins suggests a broad, cross-kingdom utility for these sequences in protein interaction networks. A notable example is found in Actinobacteria, which have acquired a eukaryotic-like 20S proteasome through horizontal gene transfer. This proteasome is regulated by a prokaryotic ubiquitin-like protein (Pup) and an ATPase cap (Mpa) that contains a variant of the HbYX motif and is structurally similar to eukaryotic AAA-ATPases [59, 60]. In Actinobacteria, we identified 982 proteins that feature the HbYX motif (Fig. 1B).

Archaeal proteomes revealed ~7.2 thousand proteins with the HbYX motif at their C-termini, corresponding to 0.93% of the proteins in these foundational life forms. This finding highlights the evolutionary continuity, as archaea, like eukaryotes, rely on the proteasome system for protein degradation. The archaeal proteasome system consists of a simpler 20S catalytic core and a hexameric unfoldase complex known as PAN. Similar to the eukaryotic 19S ATPase complex (Rpt1-6), the PAN AAA-ATPase employs mechanochemical forces to unfold and translocate target polypeptides into the 20S catalytic chamber [26–28]. Like all known 20S caps in eukaryotes and actinobacteria, PAN features a conserved C-terminal hydrophobic-tyrosine-X (HbYX) motif that

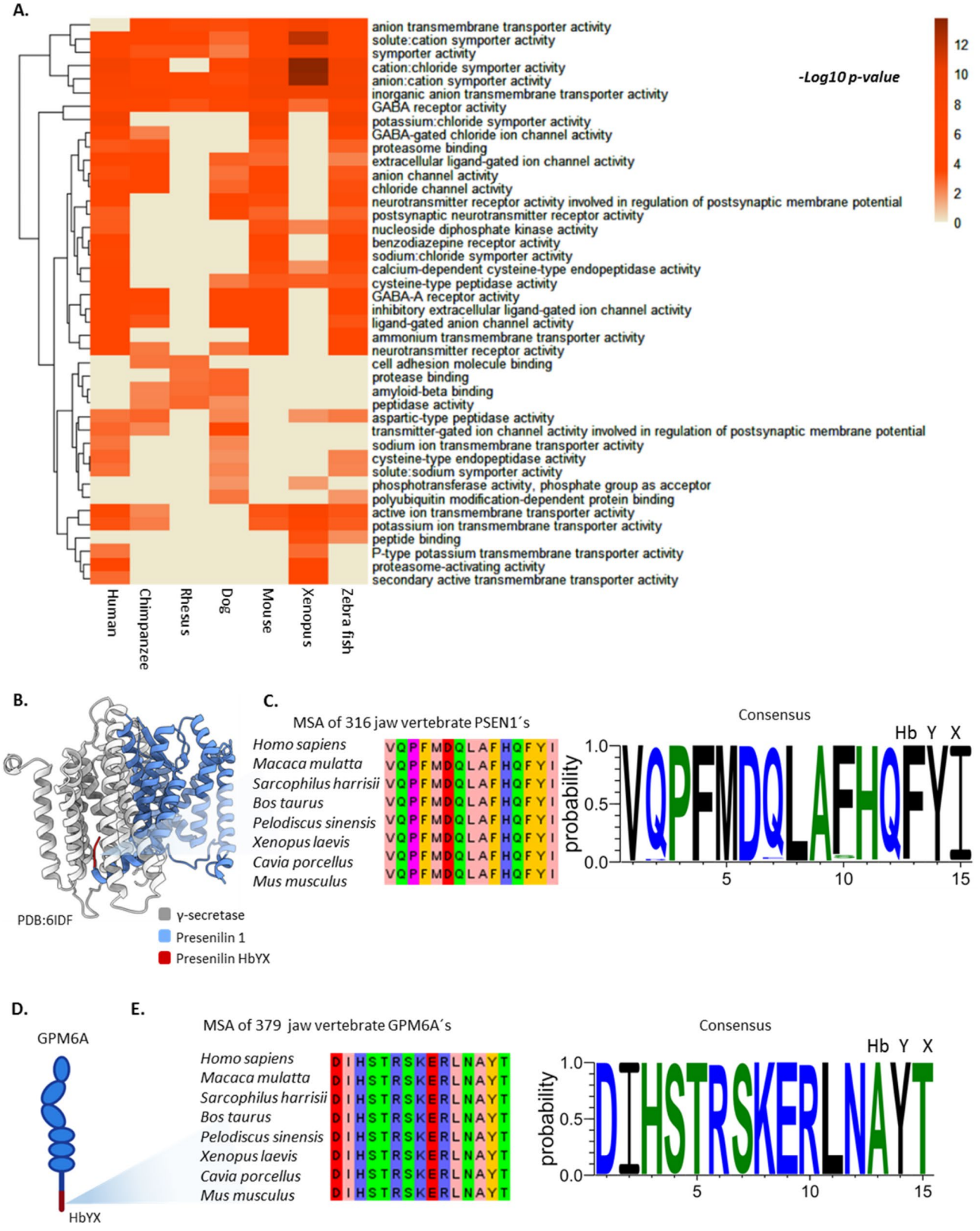

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Conserved enrichment of HbYX proteins associated with neuronal pathways among jawed vertebrates indicates conserved functions for the HbYX motif. **A.** Heatmap representing molecular function enrichment among model jawed vertebrates. The heatmap is colored by the significance of enrichment ('-log10 p-value'). Higher values indicate increased significance associated with the pathway. Rows indicate the enriched processes, and columns represent the model animals. Enrichment was determined using ClusterProfiler. Zero values indicate no enrichment found for this process among the HbYX proteins. **B.** Presenilin interaction with γ-secretase (PDB: 6IDF). This structural model represents the HbYX interaction of PSEN1 with γ-secretase, suggesting a potential binding role of the HbYX motif. γ-secretase is shown in gray, PSEN1 in blue, and the HbYX motif in red. **C.** Conservation of the HbYX motif in jawed vertebrates. A total of 316 sequences of PSEN1 were recovered from different vertebrates, and a multiple sequence alignment (MSA) was constructed using Jalview. On the right, the hidden Markov model (HMM) logo for the consensus sequence of the 316 PSEN1 sequences shows the high conservation of the HbYX motif (FYI). **D.** Representation of the neuronal membrane glycoprotein GPM6A and its HbYX motif. **E.** Conservation of the HbYX motif in jawed vertebrates A total of 379 GPM6A sequences from jawed vertebrates were retrieved, and an MSA was constructed using Jalview, highlighting the conserved features of this protein. The consensus HMM logo was then plotted, showing a highly conserved HbYX motif (AYT).

facilitates proteasome gate opening and substrate entry [26–28]. We identified 60 different PAN proteins across the archaeal proteomes (Fig-1 C). Furthermore, the VCP/p97/VAT/CDC48 AAA-ATPase, which also contains the HbYX motif, has been shown to interact with the archaeal 20S proteasome and may function as an alternative regulatory complex, although its binding to the eukaryotic 20S proteasome and its effectiveness as a proteasome cap are still under debate [30, 33, 35]. We identified 54 VAT/CDC48 proteins with the HbYX in archaeal proteomes. For perspective, the single archaea organism, *Archaeoglobus fulgidus*, has 44 HbYX proteins including one CDC48, but most of which have unknown functions. The detection of HbYX motif-containing proteins in archaea suggests a possible regulatory network conservation of activators related to this motif across these domains.

Eukaryotic organisms displayed approximately 405 thousand proteins with the HbYX motif, about 1.01% of their total protein count (Fig. 1B). This is expected given the complexity and specialization of the eukaryotic proteasome system, which needs diverse interactions with various degradation pathways. From this amount 3,266 are known activators from the 19S family of regulators and PA200 (Fig-1 C). In humans specifically, we found 234 proteins with the HbYX motif, representing 1.67% of the protein-coding sequences. This is notably lower than the anticipated random frequency of 2.5% for any given amino acid sequence, implying a potential evolutionary pressure to constrain the abundance of HbYX motif proteins, possibly due to their specialized function in proteasome regulation. Prior studies have also attempted to identify Proteasome Interacting Proteins (PIPs) by various methods such as pulldowns and mass spec. When we compared the known PIPs with all identified human HbYX proteins in ProEnd we found that 22.2% of these human HbYX proteins have indeed been identified previously as PIPs [61–63](Fig. 1D). These findings provide independent reinforcement to the idea that the HbYX proteins identified by ProEnd have a high probability to bind to the proteasome. This validation is extended to the archeal organism *Thermoplasma acidophilum*, a widely known model organism used to understand proteasome function. In this organism independent validations have shown that 25% of the HbYX containing proteins have been detected by pulldown with the 20S proteasome [30] (Fig. 1D). These independent validations show the potential of the HbYX motif as a cross-kingdom characteristic that indicates the ability to interact with the proteasome and potentializes all proteins containing HbYX in the C-terminus as PiPs, many of which may be modulators of proteasome function or related degradation pathways.

Biological-driven hypothesIs generation with ProEnd: HbYX motifs in neuronal pathways: enrichment and implications for proteasome interaction

The most well-known and studied function of the HbYX motifs is its ability to bind to and activate the 20S proteasome. Known regulators such as PA200, PI31, and PACs, are capable of activating, inhibiting, and assembling the proteasome, respectively [22, 64, 65]. To determine additional functional diversity associated with the HbYX motif identified by our approach we performed a gene enrichment analysis [44] in model gnathostomata organisms. Among jaw vertebrates the enrichment revealed a prevalence of HbYX proteins in transmembrane transport, neurotransmitter activity, and GABA receptor-related functions (Fig. 2A). In humans, from the pool of HbYX proteins, seven have been identified as involved in Alzheimer's disease pathways based on literature (not counting the proteasome itself), and all but one bind APP and include Presenillin 1 & 2 [66]. Interestingly, Presenilin's HbYX motif is extracellular, interacts with a pocket on the gamma-secretase subunit [67] (Fig. 2B), affecting its substrate gating and it is highly conserved among in jaw vertebrates (Fig. 2C).

This suggests that HbYX motifs may regulate transmembrane enzymatic function/activity and could potentially recruit extracellular proteasomes when not bound to transmembrane partners. The 20S is present in the cerebrospinal fluid, indicating putative extracellular 20S docking sites with unknown functions. Most interestingly we found that GPM6A, which has been suggested to associate with Neuronal Membrane Proteasomes (NMPs [68]) is a HbYX protein and its motif (AYT) is conserved among jawed vertebrates (Fig. 2D-E). The NMP is emerging as an important factor in regulating neuronal activity

and is proposed to be associated with neurodegeneration [68, 69]. The existence on NMP open the possibilities for the HbYX proteins found in the neuronal context such as GABA receptors and others in our database as potential recruiters or modulators of proteasome function in different cellular contexts from those expected so far, where protein degradation by the proteasome remains uncharted territory. Further validations are necessary to fully understand the context and function of these molecules and whether there is a proteasome-independent function for HbYX molecules, which seems likely, or if it is recruited to perform localized functions in cellular context. These are just a few examples of the potential hypotheses generated from our HbYX protein database, as HbYX-proteins can be found in almost every cellular compartment, such as cytosol, nucleus, ER, Golgi, nucleoplasm, mitochondria, peroxisomes, and nuclear speckles.

ProEnd identification of neurotensin-proteasome interaction: a case study in HbYX protein function

To validate the utility and usage of ProEnd for novel proteasome interacting proteins, we tested 2 examples from different organisms. The first one we selected related to neuronal regulation is an interesting example of a highly conserved protein among jawed vertebrates, neurotensin, a small protein that contains the HbYX motif in the form of YYY. This motif is conserved across all vertebrates with minimal changes in *sauropsides* where the initial Y is replaced with a S (Fig. 3A). Under physiological conditions Neurotensin is a regulatory peptide of 13 amino acids originated by the cleavage of the full-length precursor protein, predominantly located in the gut and brain influencing various dopaminergic pathways [70, 71]. The cleavage of the neurotensin precursor results in the release of neurotensin and neuromedin N, along with other peptide fragments. These peptides are less potent than neurotensin but are also less sensitive to degradation, which may allow them to act as longer-lasting modulators in various physiological situations [72, 73].

One of these fragments is the very C-terminal which contains the HbYX YYY motif. To test whether it is associated with the proteasome, human biotin- fused to neurotensin C-terminal HbYX peptide was synthesized to perform a pull-down assays on HEK293 cell lysates. Our results showed that neurotensin can pull down the 20S proteasome from human cells (Fig. 3B-C), indicating an interaction between this highly conserved HbYX-protein and proteasome degradation pathways. This suggests that highly conserved proteins with the HbYX motif are capable of binding to the proteasome, hence, its function, conserved through evolution can be related to proteasome binding and regulation. It's important to note that, to date, there is no evidence in the literature suggesting that the neurotensin peptide or other HbYX motif-containing proteins related to neurotransmission, such as GABA receptors, are associated with proteasome binding, revealing that many of the HbYX molecules identified could recruit the proteasome to execute functions in different cellular contexts, increasing the number of functions associated with the proteasome and opening the door to new possibilities for understanding the cellular context of degradation in each cell type. Neurotensin presents an example of how proteins found with HbYX motif in ProEnd database uncover a rich niche of new functions, hypotheses and approaches to understanding proteasome biology.

ProEnd discovery and functional analysis of a HbYX proteasome regulator in M. jannaschii

Our second example comes from archaea, highlighting the proteasome activities mediated by HbYX proteins in these organisms. Specifically, we investigated a protein from *Methanocaldococcus jannaschii*, this specie has the known proteasome activator PAN [74], but with ProEnd help we identified a HbYX containing protein annotated as putative 26S proteasome regulatory subunit by similarity according to UniProt (MJ1494, hereafter referred as PAN-2). Motivated by its potential role as a proteasome regulator, we generated this HbYX protein recombinantly to assess its regulatory capabilities within the proteasome. To assess its ability to bind and regulate the archaeal proteasome T20S, we employed an experimental approach similar to Fig. 3B but with slight modifications. First, recombinant T20S was purified using Ni-NTA affinity chromatography followed by TEV cleavage in order to remove the 6His tag. Concurrently, recombinant HbYX protein PAN-2 was purified through sonication and Ni-NTA affinity chromatography. Prior to elution, purified and TEV cut T20S was added to columns containing HbYX-bound PAN-2 and washed thoroughly. The HbYX protein was then eluted using wash buffer supplemented with 150mM imidazole.

The presence of T20S in the elution fractions was determined by a proteasome activity assay using the fluorogenic substrate suc-LLVY-AMC, which is cleaved by the T20S. The specificity of this assay allows us to measure the presence of T20S accurately and to calculate the initial velocity (Vo) of substrate degradation, unaffected by any regulatory effects on the T20S gate, since gate-closure does not affect LLVY degradation. We found that PAN-2 pulled down 7X more T20S, than did our control Ni-NTA columns (Fig. 4A). This result validates PAN-2 as a HbYX containing protein with the ability to bind to the proteasome, adding another validation to support the HbYX containing proteins found in our ProEnd database as proteasome regulators. Additionally, to understand PAN-2 structure conformation we predicted its

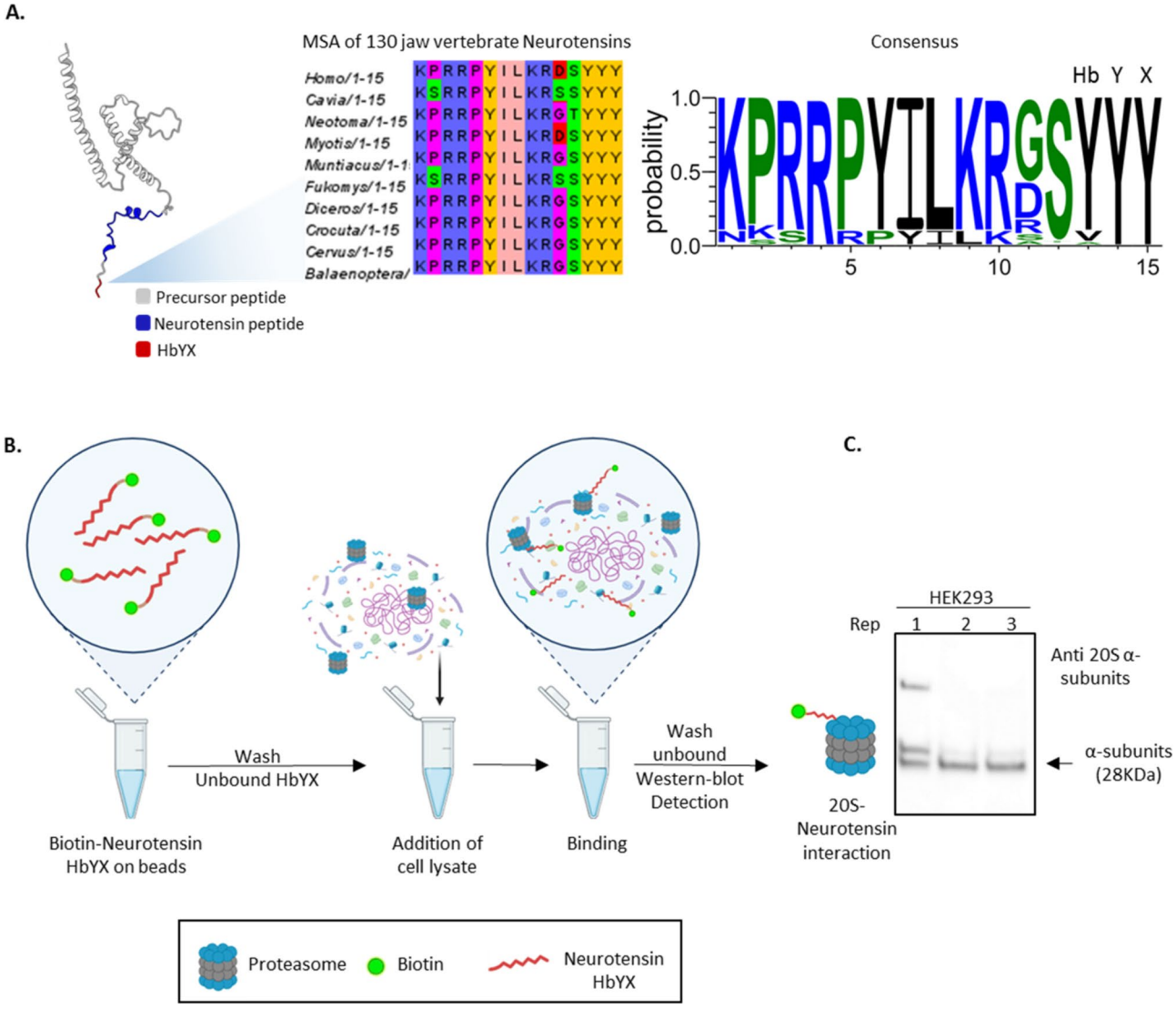

Fig. 3 Neurotensin as a conserved HbYX-containing protein with the ability to bind the proteasome. **A.** AlphaFold prediction of the neurotensin precursor (AAF-P30990-F1) showing the Neurotensin peptide in blue and the HbYX motif in red. For the HbYX motif, a multiple sequence alignment (MSA) was constructed using neurotensin sequences from 130 vertebrates to show conservation of the HbYX motif. The consensus for the HbYX motif was constructed using an HMM logo, revealing a highly conserved YYY motif, with occasional occurrences of valine and alanine in the hydrophobic position. **B.** Validation pull-down assay performed with the Neurotensin HbYX peptide fused to biotin. Briefly, biotinylated Neurotensin was incubated with avidin beads. After washing off unbound HbYX, cell lysates from HEK293 cells were added for binding. Following incubation, unbound proteins were washed off, and detection of the proteasome was performed through western blotting. **C.** Western blot for the pull-down assay described in B, probing for alpha subunits with an expected molecular weight of 28 kDa. Three different replicates of HEK293 cell pull-downs were tested, showing the presence of alpha subunit proteasome bands. This indicates that the neurotensin HbYX peptide can pull down proteasomes, demonstrating that neurotensin is a proteasome interactor protein. The upper band in replicate 1 indicates the presence of the 26S proteasome

homo-oligomerization using AlphaFold Multimer [48]. We expected homo-oligomerization since PAN-2 contains an ATP binding domain in the 161–168 positions, and it is expected, based on homology with ATPases such as PAN to form hexamers. The AlphaFold prediction suggested a hexameric assembly with a pLDDT score of 0.78 (Fig. 4B and D HbYX in orange), indicating good overall structural accuracy. The Predicted Aligned Error (PAE) plot revealed high confidence in the intra-monomeric contacts and moderate confidence in the inter-monomeric interactions, consistent with the hexameric structure (Fig. 4C). While the local structures are reliable, some inter-chain interactions may require further experimental validation. According to the AlphaFold prediction the HbYX motif seems to be buried in the ATPase ring which is also a common feature shared with PAN when it is not bound to ATP and not bound to the proteasome, indicating structural similarities between PAN-1 and PAN-2. Noteworthy is that PAN-2 homology percentage to *M.jannaschii*, PAN-1 is 29.77%

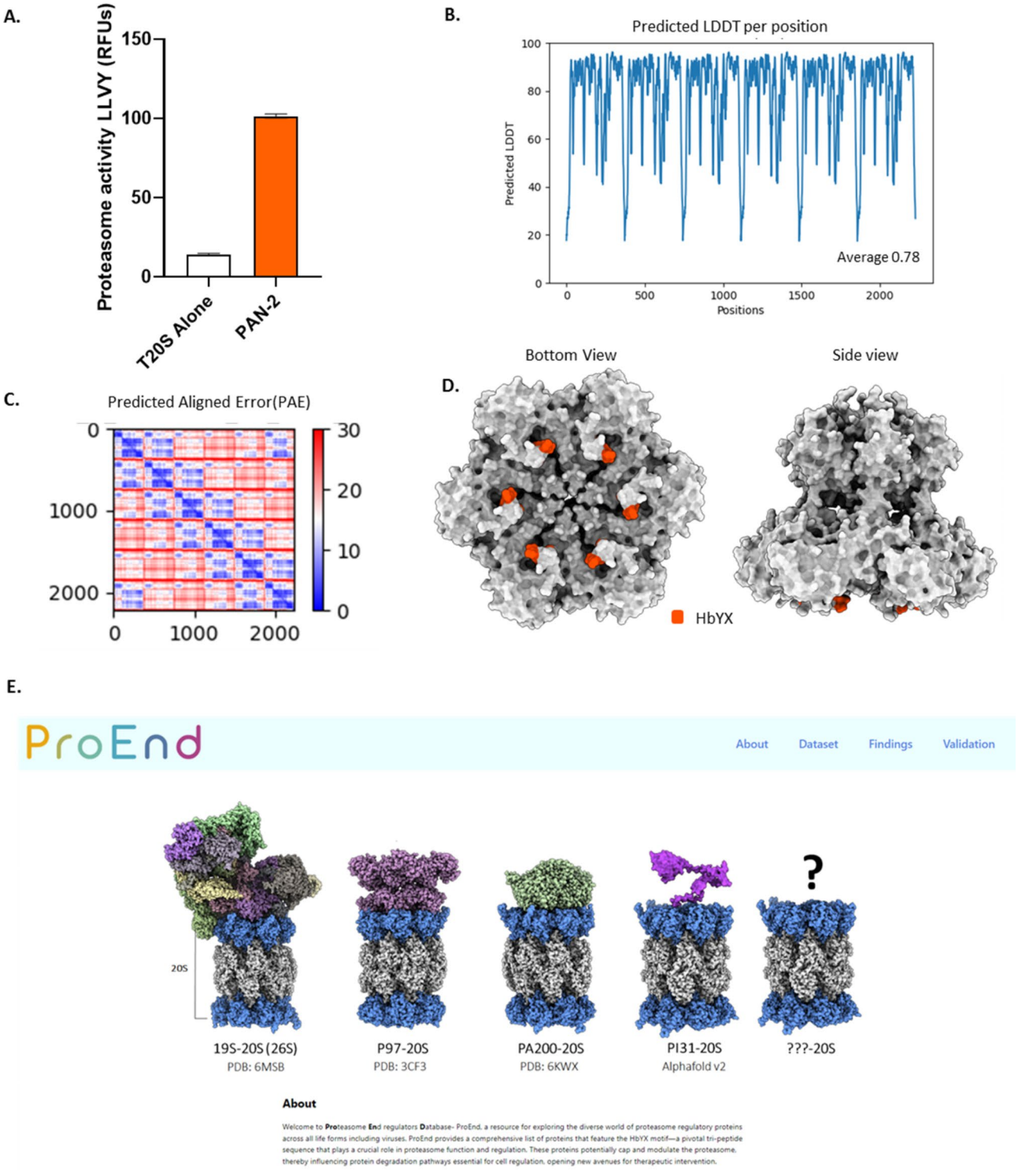

Fig. 4 (See legend on next page.)

using the BLOSUM62 matrix. While this percentage suggests some sequence similarity, it is relatively low, indicating a distant evolutionary relationship and potentially different regulatory roles or specificities. Importantly, PAN-2 does possess the N-terminal OB domain, akin to other proteasome ATPases, suggesting a conservation of structural features despite the low homology. This finding enhances our knowledge of proteasome activators in ancient organisms like archaea, demonstrating the variability in PAN-like proteins across distinct organism

(See figure on previous page.)

Fig. 4 Archaeal HbYX ATPase PAN-2 (MJ1494) binds and activates the archaeal proteasome T20S. **(A)** Rate of substrate degradation of fluorogenic LLVY after pulldown with T20S on PAN-2, pulldown with proteasome T20S alone. Stimulation of degradation was measured by the increase in LLVY hydrolysis (rfu). **(B)** AlphaFold multimer prediction LDDT indicating overall confidence in the predicted structure. Scores above 0.51 are generally accepted. Scores above 0.70 are considered to be of good quality, suggesting that most regions of the protein are predicted with high accuracy. **(C)** Predicted Aligned Error (PAE) provides insights into the accuracy of the predicted inter-residue distances. The diagonal blocks with lower PAE values (indicated by blue regions) suggest that the intra-monomeric contacts are predicted with high confidence. This aligns with the high pLDDT score (0.78), indicating reliable local structure predictions within each monomer. The off-diagonal blocks show higher PAE values (indicated by red regions), reflecting moderate confidence in the inter-monomeric contacts, indicating some uncertainty in the predicted interactions between monomers. The repeating pattern of the blocks in the PAE plot is consistent with the hexameric nature of the protein, suggesting that each monomer's predicted structure is similar and forms repetitive interactions within the hexamer **(D)** Bottom and side views of the structural representation predicted by AlphaFold multimer visualized with ChimeraX. The complex is shown in gray with the HbYX motif highlighted in orange for visualization. The HbYX motif appears to be buried inside the ATPase ring, a common feature shared with PAN when not engaged with the 20S proteasome. **(E)** Screenshot of the main page of ProEnd. The database is available at http://proend.org/

families and it also illustrates the evolutionary adaptation and functional differentiation of proteasome regulators in diverse biological contexts.

A web portal for HbYX motif-containing proteins

To allow researchers to interact with the comprehensive dataset of HbYX motif-containing proteins, we created ProEnd, a user-friendly online database (first released in 2024 Fig. 4E). Users can explore an expandable, interactive matrix or a sortable list that can be filtered by species, domain, or gene name. Clicking on a matrix tile or list entry displays an information page that includes an interactive protein structure viewer, where AlphaFold structure predictions of the HbYX motifs can be examined and downloaded. The information page also provides UniProt entry information. These features facilitate rapid searching, visualization, and analysis of several HbYX proteins, making ProEnd an invaluable resource for the global scientific community. The database is hosted on Amazon Web Services using MongoDB, ensuring high availability and performance. Researchers can access ProEnd database at http://proend.org/ and navigate the data to support their studies in proteasome regulation and beyond. Additionally, the ProEnd tool is available on GitHub at https://github.com/drTakuOmics/proend-scripts, where a concise tutorial explains how to perform searches of your data.

Discussion

The ProEnd database addresses a critical gap in proteasome research by providing a targeted resource for exploring proteins with the HbYX motif, which are known to influence proteasome activity either through activation or inhibition. This motif is critically important due to its therapeutic potential and its potential role in regulating specific cellular pathways through the 20S proteasome. Previously, the research community lacked a comprehensive source for identifying and generating hypothesis with these potential proteasome interactors, a gap now filled by ProEnd.

Our results highlight the evolutionary prevalence of the HbYX motif not only in well-established domains such as eukaryotes, actinobacteria, and archaea, which possess the 20S proteasome, but also in viruses. This suggests evolutionary pressures might be shaping the development of this motif in viruses to either evade or hijack the host's degradation system. This is a very interesting question which needs further exploration but indicates this motif as a gold standard for protein degradation modulation across all organisms. A critical component to support the potential of the HbYX motif are 20S pulldown validations where at least 25% of the HbYX-containing proteins in human do indeed interact with the proteasome, a result corroborated independently by other researchers. This confirms the relevance of HbYX-containing proteins as a source of potential proteasome interactors.

Furthermore, the enrichment of HbYX proteins in specific processes such as ion channels and neurotransmission open new avenues for understanding the functions of these motifs beyond proteasome interaction. This is particularly intriguing in the context of the neuronal membrane proteasome (NMP), where the full range of interactors remains undefined. Additionally, specific pipelines will be needed to understand the potential role of the 20S association with these new HbYX proteins. In addition, our validation of neurotensin as a proteasome interactor exemplifies the utility of ProEnd. Neurotensin, a neurotransmitter highly conserved, was shown to interact with the 20S proteasome, suggesting potential involvement in neurotransmission pathways through the presence of the HbYX motif. While this does not imply that neurotensin induces proteasome activity or that this interaction is its primary function, the interaction suggests a potential role and raises important questions about potential functional associations in neuronal contexts. Further research, along with the development of specific analytical pipelines, will be essential to explore the implications of this 20S association more comprehensively.

Additionally, the validation of PAN-2, a new PAN-like molecule from the archaeal organism *Methanocaldococcus jannaschii*, which binds to and activates the archaeal 20S proteasome, illustrates another significant

application of the ProEnd database. This protein, identified and modeled using AlphaFold Multimer, highlights the potential of ProEnd to generate hypotheses and guide new questions in proteasome biology. Although further characterization is required to understand the structural specifics and substrates of this new activator, this finding demonstrates the valuable role of our database in advancing proteasome research across various organisms.

Conclusions

The establishment of the ProEnd database marks a significant advancement in the field of proteasome biology, providing an invaluable resource for identifying and characterizing HbYX motif-containing proteins across a broad spectrum of species. By offering a comprehensive curated catalog and an interactive platform, ProEnd facilitates the exploration of potential proteasomal regulators and their evolutionary conservation, uncovering their potential diverse roles in cellular processes. In this study, we identified a conserved enrichment of HbYX motif-containing proteins in jaw vertebrates associated with neuronal pathways (Fig. 2), suggesting that the HbYX motif may regulate proteasome function in unexpected contexts such as cellular membranes or in extracellular recruitment of proteasomes within neuronal environments. This finding aligns with emerging trends in research focusing on the neuronal membrane proteasome (NMP). We further validated the involvement of two novel HbYX-containing proteins. Firstly, neurotensin, a well-known neurotransmitter, was demonstrated to interact with the 20S proteasome in human cell lines through pull-down experiments (Fig. 3), highlighting its potential role in proteasome modulation. Secondly, our experimental results reveal that the PAN-like, HbYX-containing protein PAN-2 from the archaeal organism *M.jannaschii* interacts with and activates the archaeal T20S proteasome (Fig. 4), illustrating the potential for similar molecular mechanisms across other archaeal species. Future directions will focus on optimizing and validating our dataset and consistently applying our methodologies to further elucidate the biological functions associated with HbYX-containing proteins. The findings reported here highlight the prevalence and importance of HbYX motifs in proteasome regulation, with implications extending from basic biological research to potential therapeutic applications. This database is expected to not only enhance our understanding of proteasome mechanisms but also opens new avenues for the discovery of novel proteasome modulators, with potential far-reaching impacts to understanding proteasome biology and treating human disease.

In conclusion, the ProEnd database is proving to be a vital resource for the scientific community, enabling a comprehensive exploration of proteasome regulation biology and facilitating discoveries that could have far-reaching implications for understanding and treating diseases linked to proteasome dysfunction.

Acknowledgements
We extend our appreciation to the members of the Smith Lab for their invaluable discussions and feedback, which significantly enriched the development of this manuscript, particularly to Dr. Chuah for her expert insights. We also owe a special thanks to Dr. Chris Adami and the team at the Institute for Cyber-Enabled Research at Michigan State University. Their computational resources and services were instrumental in enabling the AlphaFold oligomer modelling. Lastly, we are grateful for the financial support from the NIH, specifically grants R01AG064188 and R01GM107129 awarded to D.M.S., which made this work possible.

Author contributions
D.S.T. and A.G.D. developed the AWK scripts for data retrieving and processing, C.M. and D.S.T created the database and website. D.S.T performed the sequence alignments, enrichment and HMM analyses. J.S performed cell culture, neurotensin pulldowns and western blots. G.D.H. performed the purification pulldowns and activity assay of MJ1494 protein. A.G.D performed the alphafold multimer reconstructions and D.S.T processed the results and structures. Results were analyzed and interpreted by D.S.T., G.D.H., A.G.D. and D.M.S. Manuscript was prepared by D.S.T., G.D.H. and D.M.S. All authors reviewed the results and approved the final version of this manuscript.

Funding
Financial support from the NIH, specifically grants R01AG064188 and R01GM107129 awarded to D.M.S., which made this work possible.

Data availability
All data supporting the findings of this study are available within the paper and its Supplementary Information. The related Proend database is publicly available at proend.org.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 29 July 2024 / Accepted: 3 October 2024
Published online: 13 October 2024

References

1. Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. Physiol Rev. 2002;82:373–428. https://doi.org/10.1152/PHYSREV.00027.2001/ASSET/IMAGES/LARGE/9J0220190106.JPEG.
2. Hoyt MA, Coffino P. Ubiquitin-proteasome system. Cell Mol Life Sci C 2004 6113. 2004;61:1596–600. https://doi.org/10.1007/S00018-004-4133-9
3. Deshmukh FK, Yaffe D, Olshina MA, Ben-Nissan G, Sharon M. The Contribution of the 20S Proteasome to Proteostasis. Biomol 2019;9;190.2019;9:190. https://doi.org/10.3390/BIOM9050190
4. Spataro V, Norbury C, Harris AL. The ubiquitin-proteasome pathway in cancer. Br J Cancer 1998 773. 1998;77:448–55. https://doi.org/10.1038/bjc.1998.71
5. 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;17:351–65. https://doi.org/10.1016/J.MOLCEL.2004.12.021.
6. Tai HC, Schuman EM. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. Nat Rev Neurosci 2008 911. 2008;9:826–38. https://doi.org/10.1038/nrn2499.

7. Schmidt MF, Gan ZY, Komander D, Dewson G. Ubiquitin signalling in neurodegeneration: mechanisms and therapeutic opportunities. Cell Death Differ 2021 282. 2021;28:570–90. https://doi.org/10.1038/s41418-020-00706-7.
8. Ciechanover A, Kwon YT. Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. Exp Mol Med 2015 473. 2015;47:e147–147. https://doi.org/10.1038/emm.2014.117.
9. Whittemore SR, Saraswat Ohri S, Forston MD, Wei GZ, Hetman M. The Proteostasis Network: A Global Therapeutic Target for Neuroprotection after Spinal Cord Injury. Cells 2022;11:3339.2022;11:3339. https://doi.org/10.3390/CELLS11213339
10. Kane RC, Bross PF, Farrell AT, Pazdur R, Velcade®:. U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. Oncologist. 2003;8:508–13. https://doi.org/10.1634/THEONCOLOGIST.8-6-508.
11. Kane RC, Dagher R, Farrell A, Ko CW, Sridhara R, Justice R, et al. Bortezomib for the Treatment of Mantle Cell Lymphoma. Clin Cancer Res. 2007;13:5291–4. https://doi.org/10.1158/1078-0432.CCR-07-0871.
12. Richardson PG, Mitsiades C, Hideshima T, Anderson KC, Bortezomib. Proteasome inhibition as an effective anticancer therapy. Annu Rev Med. 2006;57 Volume 57:2006:33–47. https://doi.org/10.1146/ANNUREV.MED.57.042905.122625/CITE/REFWORKS
13. Borissenko L, Groll M. 20S proteasome and its inhibitors: crystallographic knowledge for drug development. Chem Rev. 2007;107:687–717. https://doi.org/10.1021/CR0502504/ASSET/IMAGES/LARGE/CR0502504F00008.JPEG.
14. Thibaudeau TA, Anderson RT, Smith DM. A common mechanism of proteasome impairment by neurodegenerative disease-associated oligomers. Nat Commun 2018 91. 2018;9:1–14. https://doi.org/10.1038/s41467-018-03509-0.
15. Coux O, Tanaka K, Goldberg AL, Structure and functions of, the 20S AND 26S proteasomes. https://doi.org/101146/annurev.bi65070196004101. 2003;65:801–47. https://doi.org/10.1146/ANNUREV.BI.65.070196.004101
16. Groll M, Huber R. Substrate access and processing by the 20S proteasome core particle. Int J Biochem Cell Biol. 2003;35:606–16. https://doi.org/10.1016/S1357-2725(02)00390-4.
17. Groll M, Bajorek M, Köhler A, Moroder L, Rubin DM, Huber R, et al. A gated channel into the proteasome core particle. Nat Struct Biol 2000 711. 2000;7:1062–7. https://doi.org/10.1038/80992.
18. Groll M, Ditzel L, Löwe J, Stock D, Bochtler M, Bartunik HD, et al. Structure of 20S proteasome from yeast at 2.4Å resolution. Nat 1997 3866624. 1997;386:463–71. https://doi.org/10.1038/386463a0.
19. Förster A, Whitby FG, Hill CP. The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. EMBO J. 2003;22:4356–64. https://doi.org/10.1093/EMBOJ/CDG436.
20. 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;27:731–44. https://doi.org/10.1016/j.molcel.2007.06.033.
21. Chu-Ping M, Willy PJ, Slaughter CA, DeMartino GN. PA28, an activator of the 20 S proteasome, is inactivated by proteolytic modification at its carboxyl terminus. J Biol Chem. 1993;268:22514–9. https://doi.org/10.1016/s0021-9258(18)41559-1.
22. Toste Rêgo A, da Fonseca PCA. Characterization of fully recombinant human 20S and 20S-PA200 Proteasome complexes. Mol Cell. 2019;76:138–e1475. https://doi.org/10.1016/j.molcel.2019.07.014.
23. Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC, et al. Structural basis for the activation of 20S proteasomes by 11S regulators. Nat 2000 4086808. 2000;408:115–20. https://doi.org/10.1038/35040607.
24. Raule M, Cerruti F, Benaroudj N, Migotti R, Kikuchi J, Bachi A, et al. PA28αβ reduces size and increases hydrophilicity of 20S immunoproteasome peptide products. Chem Biol. 2014;21:470–80.
25. Fort P, Kajava AV, Delsuc F, Coux O. Evolution of Proteasome regulators in eukaryotes. Genome Biol Evol. 2015;7:1363–79. https://doi.org/10.1093/GBE/EVV068.
26. Benaroudj N, Zwickl P, Seemüller E, Baumeister W, Goldberg AL. ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. Mol Cell. 2003;11:69–78. https://doi.org/10.1016/S1097-2765(02)00775-X.
27. Smith DM, Kafri G, Cheng Y, Ng D, Walz T, Goldberg AL. ATP binding to PAN or the 26S ATPases Causes Association with the 20S proteasome, gate opening, and translocation of unfolded proteins. Mol Cell. 2005;20:687–98.
28. Majumder P, Rudack T, Beck F, Danev R, Pfeifer G, Nagy I, et al. Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. Proc Natl Acad Sci U S A. 2019;116:534–9. https://doi.org/10.1073/PNAS.1817752116/SUPPL_FILE/PNAS.1817752116.SM02.MPG.
29. Barthelme D, Chen JZ, Grabenstatter J, Baker TA, Sauer RT. Architecture and assembly of the archaeal Cdc48·20S proteasome. Proc Natl Acad Sci U S A. 2014;111:E1687–94. https://doi.org/10.1073/PNAS.1404823111/SUPPL_FILE/PNAS.201404823SI.PDF.
30. Forouzan D, Ammelburg M, Hobel CF, Ströh LJ, Sessler N, Martin J, et al. The archaeal proteasome is regulated by a network of AAA ATPases. J Biol Chem. 2012;287:39254–62.
31. Zwickl P, Ng D, Woo KM, Klenk HP, Goldberg AL. An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. J Biol Chem. 1999;274:26008–14. https://doi.org/10.1074/jbc.274.37.26008.
32. 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;18:589–99. https://doi.org/10.1016/j.molcel.2005.04.016.
33. Braxton JR, Southworth DR. Structural insights of the p97/VCP AAA + ATPase: how adapter interactions coordinate diverse cellular functionality. J Biol Chem. 2023;299:105182.
34. van den Boom J, Meyer H. VCP/p97-Mediated unfolding as a Principle in protein homeostasis and signaling. Mol Cell. 2018;69:182–94.
35. Gerega A, Rockel B, Peters J, Tamura T, Baumeister W, Zwickl P. VAT, the Thermoplasma homolog of mammalian p97/VCP, is an N domain-regulated protein unfoldase. J Biol Chem. 2005;280:42856–62. https://doi.org/10.1074/jbc.M510592200.
36. Opoku-Nsiah KA, de la Pena AH, Williams SK, Chopra N, Sali A, Lander GC, et al. The YΦ motif defines the structure-activity relationships of human 20S proteasome activators. Nat Commun 2022 131. 2022;13:1–12. https://doi.org/10.1038/s41467-022-28864-x.
37. Chuah JJY, Thibaudeau TA, Smith DM. Minimal mechanistic component of HbYX-dependent proteasome activation that reverses impairment by neurodegenerative-associated oligomers. Commun Biol 2023 61. 2023;6:1–10. https://doi.org/10.1038/s42003-023-05082-9.
38. Chuah JJY, Rexroad MS, Smith DM. High resolution structures define divergent and convergent mechanisms of archaeal proteasome activation. Commun Biol 2023 61. 2023;6:1–16. https://doi.org/10.1038/s42003-023-05123-3.
39. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9. https://doi.org/10.1093/NAR/GKAA913.
40. Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic Acids Res. 2013;41:D344–7. https://doi.org/10.1093/NAR/GKS1067.
41. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43:W39–49. https://doi.org/10.1093/NAR/GKV416.
42. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32 suppl_2:W20–5. https://doi.org/10.1093/NAR/GKH435.
43. Gardner PP, Paterson JM, McGimpsey S, Ashari-Ghomi F, Umu SU, Pawlik A, et al. Sustained software development, not number of citations or journal choice, is indicative of accurate bioinformatic software. Genome Biol. 2022;23:1–13. https://doi.org/10.1186/S13059-022-02625-X/FIGURES/2.
44. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov. 2021;2:100141.
45. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91. https://doi.org/10.1093/BIOINFORMATICS/BTP033.
46. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence Logo Generator. Genome Res. 2004;14:1188–90. https://doi.org/10.1101/GR.849004.
47. Seemüller E, Lupas A, Stock D, Löwe J, Huber R, Baumeister W. Proteasome from Thermoplasma acidophilum: a threonine protease. Science. 1995;268:579–82. https://doi.org/10.1126/SCIENCE.7725107.
48. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022https://doi.org/10.1101/2021.10.04.463034.
49. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the Universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31. https://doi.org/10.1093/NAR/GKAC1052.
50. Sijts EJAM, Kloetzel PM. The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci. 2011;68:1491–502. https://doi.org/10.1007/S00018-011-0657-Y/METRICS.
51. Rock KL, Goldberg AL. Degradation of cell proteins and the generation of MHC class I-presented peptides. Annu Rev Immunol. 1999;17 Volume 17,

1999:739–79. https://doi.org/10.1146/ANNUREV.IMMUNOL.17.1.739/CITE/REFWORKS
52. Luo H. Interplay between the virus and the ubiquitin–proteasome system: molecular mechanism of viral pathogenesis. Curr Opin Virol. 2016;17:1–10.
53. Gottesman S, Roche E, Zhou YN, Sauer RT. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev. 1998;12:1338. https://doi.org/10.1101/GAD.12.9.1338.
54. Kress W, Maglica Ž, Weber-Ban E. Clp chaperone–proteases: structure and function. Res Microbiol. 2009;160:618–28.
55. Ruiz-González MX, Marín I. Proteasome-related HslU and HslV genes typical of eubacteria are widespread in eukaryotes. J Mol Evol. 2006;63:504–12. https://doi.org/10.1007/S00239-005-0282-1/FIGURES/5.
56. Rohrwild M, Coux O, Huang HC, Moerschell RP, Yoo SJ, Seol JH, et al. HslV-HslU: a novel ATP-dependent protease complex in Escherichia coli related to the eukaryotic proteasome. Proc Natl Acad Sci U S A. 1996;93:5808. https://doi.org/10.1073/PNAS.93.12.5808.
57. Bouchnak I, van Wijk KJ. Structure, function, and substrates of Clp AAA + protease systems in cyanobacteria, plastids, and apicoplasts: a comparative analysis. J Biol Chem. 2021;296:100338. https://doi.org/10.1016/J.JBC.2021.100338.
58. Trentini DB, Suskiewicz MJ, Heuck A, Kurzbauer R, Deszcz L, Mechtler K, et al. Arginine phosphorylation marks proteins for degradation by a Clp protease. Nat 2016 5397627. 2016;539:48–53. https://doi.org/10.1038/nature20122.
59. Darwin KH, Ehrt S, Gutierrez-Ramos JC, Weich N, Nathan CF. The Proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. Sci (80-). 2003;302:1963–6. https://doi.org/10.1126/SCIENCE.1091176/SUPPL_FILE/DARWIN.SOM.PDF.
60. Kavalchuk M, Jomaa A, Müller AU, Weber-Ban E. Structural basis of prokaryotic ubiquitin-like protein engagement and translocation by the mycobacterial mpa-proteasome complex. Nat Commun 2022 131. 2022;13:1–13. https://doi.org/10.1038/s41467-021-27787-3.
61. Pepelnjak M, Rogawski R, Arkind G, Leushkin Y, Fainer I, Ben-Nissan G, et al. Systematic identification of 20S proteasome substrates. Mol Syst Biol. 2024. https://doi.org/10.1038/S44320-024-00015-Y/SUPPL_FILE/44320_2024_15_MOESM10_ESM.PDF.
62. Fabre B, Lambour T, Garrigues L, Amalric F, Vigneron N, Menneteau T, et al. Deciphering preferential interactions within supramolecular protein complexes: the proteasome case. Mol Syst Biol. 2015;11:771. https://doi.org/10.15252/MSB.20145497/SUPPL_FILE/MSB145497.REVIEWER_COMMENTS.PDF.
63. Fabre B, Lambour T, Delobel J, Amalric F, Monsarrat B, Burlet-Schiltz O, et al. Subcellular distribution and Dynamics of active Proteasome complexes unraveled by a Workflow combining in vivo complex cross-linking and quantitative proteomics. Mol Cell Proteom. 2013;12:687–99.
64. Hsu HC, Wang J, Kjellgren A, Li H, DeMartino GN. High-resolution structure of mammalian PI31–20S proteasome complex reveals mechanism of proteasome inhibition. J Biol Chem. 2023;299:104862.
65. Adolf F, Du J, Goodall EA, Walsh RM, Rawson S, von Gronau S, et al. Visualizing chaperone-mediated multistep assembly of the human 20S proteasome. Nat Struct Mol Biol 2024. 2024;1–13. https://doi.org/10.1038/s41594-024-01268-9.
66. Da Costa CA, Ancolio K, Checler F. C-terminal maturation fragments of presenilin 1 and 2 control secretion of APPα and Aβ by human cells and are degraded by proteasome. Mol Med. 1999;5:160–8. https://doi.org/10.1007/BF03402059/FIGURES/4.
67. Yang G, Zhou R, Zhou Q, Guo X, Yan C, Ke M, et al. Structural basis of notch recognition by human γ-secretase. Nat 2018 5657738. 2018;565:192–7. https://doi.org/10.1038/s41586-018-0813-8.
68. Ramachandran KV, Margolis SS. A mammalian nervous-system-specific plasma membrane proteasome complex that modulates neuronal function. Nat Struct Mol Biol 2017 244. 2017;24:419–30. https://doi.org/10.1038/nsmb.3389.
69. Eric Villaló Landeros A, Kho SC, Church TR, Delannoy M, Caterina MJ, Margolis SS, et al. The nociceptive activity of peripheral sensory neurons is modulated by the neuronal membrane proteasome. Cell Rep. 2024;43:114058. https://doi.org/10.1016/J.CELREP.2024.114058.
70. Christou N, Blondy S, David V, Verdier M, Lalloué F, Jauberteau MO, et al. Neurotensin pathway in digestive cancers and clinical applications: an overview. Cell Death Dis 2020 1112. 2020;11:1–12. https://doi.org/10.1038/s41419-020-03245-8.
71. Pomrenze MB, Giovanetti SM, Maiya R, Gordon AG, Kreeger LJ, Messing RO. Dissecting the roles of GABA and neuropeptides from Rat Central Amygdala CRF neurons in anxiety and fear learning. Cell Rep. 2019;29:13–e214. https://doi.org/10.1016/j.celrep.2019.08.083.
72. Carraway RE, Mitra SP. Differential processing of neurotensin/neuromedin N precursor(s) in canine brain and intestine. J Biol Chem. 1990;265:8627–31. https://doi.org/10.1016/s0021-9258(19)38933-1.
73. Kitabgi P. Neurotensin and neuromedin N are differentially processed from a common precursor by prohormone convertases in tissues and cell lines. Results Probl Cell Differ. 2010;50:85–96. https://doi.org/10.1007/400_2009_27/TABLES/1.
74. Wilson HL, Ou MS, Aldrich HC, Maupin-Furlow J. Biochemical and physical properties of the Methanococcus jannaschii 20S proteasome and PAN, a Homolog of the ATPase (rpt) subunits of the Eucaryal 26S proteasome. J Bacteriol. 2000;182:1680. https://doi.org/10.1128/JB.182.6.1680-1692.2000.

Publisher's note