Mutations in α-synuclein, TDP-43 and tau prolong protein half-life through diminished degradation by lysosomal proteases

Aimee Kao (Aimee.Kao@ucsf.edu)
UCSF https://orcid.org/0000-0002-7686-7968
Paul Sampognaro
University of California, San Francisco
Shruti Arya
University of California, San Francisco https://orcid.org/0000-0001-5787-3927
Giselle Knudsen
Alaunus Biosciences
Emma Gunderson
University of California, San Francisco
Angelica Sandoval-Perez
University of California, San Francisco
Charles Craik
University of California San Francisco https://orcid.org/0000-0001-7704-9185
Matthew Jacobson
UCSF https://orcid.org/0000-0001-6262-655X

Article

Keywords:

Posted Date: March 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1451319/v1

Abstract

Autosomal dominant mutations in α-synuclein, TDP-43 and tau promote protein aggregation and neurodegeneration. Rates of aggregation are highly dependent on protein concentration, which can be regulated via lysosomal proteolysis. Lysosomal cathepsins recognize specific linear amino acid sequences; thus, mutations in α-synuclein, TDP-43 and tau have the potential to impact protein half-life by impairing lysosomal degradation. To test this, we first generated comprehensive proteolysis maps containing cathepsin cleavage sites for each of these disease-associated proteins. In silico analysis of these maps suggested that certain mutations would diminish cathepsin cleavage, a prediction we validated utilizing in vitro protease assays. We further demonstrated that lysosomes degrade inducibly-expressed mutant forms of α-synuclein, TDP-43 and tau less efficiently than wild-type protein. Together, this study provides evidence that pathogenic mutations in α-synuclein, TDP-43 and tau directly impair their own lysosomal degradation, altering protein homeostasis by increasing the half-life of these proteins.

Introduction

Despite symptomatic and anatomical differences, neurodegenerative diseases such as Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are all characterized by the neuropathological finding of proteinaceous intraneuronal aggregates on which disease nosology is based and genetic cohorts have been identified[1]. The mechanistic basis for aggregate formation represents an area of active investigation, with recent cryo-EM studies revealing distinctive structural features of α-synuclein (α-syn), TDP-43 and tau inclusions[2–4]. Rates of protein self-association and aggregation depend sensitively on the steady-state levels present within a cell, which can rise as part of aberrant protein homeostasis[5,6]. Mutations in the genes encoding α-syn, TDP-43 and tau increase the risk for their respective diseases, and many are thought to do so by increasing the rate of protein self-association[10–12]. However, only a subset of α-syn, TDP-43 and tau mutations have been experimentally shown to increase aggregation potential while many other autosomal dominant disease mutations do not directly affect protein oligomerization[13–15]. Thus, these remaining autosomal dominant mutations in α-syn, TDP-43 and tau may increase their own steady-state levels over time via an alternative mechanism.

Lysosomes and the lysosome-dependent autophagic processes play major roles in cellular protein homeostasis[16–19]. Lysosomal cathepsins can be classified as cysteine, serine, or aspartyl proteases based on the key catalytic amino acid present in the active site. Cathepsins exhibit selectivity for their substrate clients through the recognition of specific linear sequences of amino acids. Modest pH alterations or single residue changes in the target sequence of substrate proteins can dramatically alter cleavage kinetics[7–9,20]. Although α-syn, TDP-43 and tau are all subject to lysosomal clearance [21–23], the ability of cathepsins to cleave wild type (WT) versus mutant versions of these proteins has not been comprehensively explored.

In this study, we investigated the normal lysosomal clearance of α-syn, TDP-43 and tau and found that a subset of disease mutations extended protein half-life and increased cellular protein levels. These insights arose following the generation and analysis of comprehensive maps of protease cleavage of these proteins. The maps revealed that areas dense with cleavage sites often overlapped with areas where disease mutations are concentrated. They also predicted mutations that were likely to interfere with protease cleavage. Systematic testing of a subset of these “disruptive” mutations validated these predictions both *in vitro* and in a neuronal model, the latter of which demonstrated alterations in α-syn, TDP-43 and tau half-life. Together, this study offers an additional pathobiological mechanism by which both sporadic and genetic forms of neurodegenerative disease may develop and implicates lysosomal proteases as therapeutically targetable contributors in neurodegeneration.

Results

Lysosomal proteases digest recombinant human α-syn, TDP-43 and tau in a selective fashion

Although prior studies have demonstrated that α-syn, TDP-43 and tau can be shuttled to the lysosome for degradation[21–23], the specific proteases that carry out their degradation once inside the lysosomal compartment have not been fully characterized. Thus, we set out to broadly identify which lysosomal proteases are able to cleave α-syn, TDP-43 and tau via *in vitro* cleavage assays (Fig. 1). As different cathepsins have distinct pH optima, we performed individual reactions of full-length, recombinant protein α-syn, TDP-43 or tau with each protease across a range of pH settings (3.4–7.4).

We first assayed the largest class of lysosomal proteases, the cysteine protease family[24]. Under our experimental conditions, cathepsins B (CTSB), L (CTSL), K (CTSK), S (CTSS), and V (CTSV) were capable of cleaving all three full-length proteins at or near their optimal pH within the span of 1 hour. Cathepsin C (CTSC) and asparaginyl endopeptidase (AEP) appeared to cleave α-syn and tau moderately but not TDP-43. Interestingly, cathepsin F (CTSF) could only cleave tau. Most cysteine proteases cleaved these substrates at their known pH optima (typically between pH 4.5 to 6.5). However, CTSV, an enzyme previously observed to cleave at lower pHs than other cysteine proteases[25], appeared most active against these full-length protein substrates at pH 3.4. Intriguingly, these results suggest that CTSV *in vivo* may only be active against these substrates in the context of mature, highly acidified lysosomes.

Next, we tested the serine proteases, cathepsin A (CTSA) and cathepsin G (CTSG). While CTSG was able to digest all three substrates at more neutral pHs, CTSA did not demonstrate any clear protease activity against any of our substrates in this assay. Lastly, we tested the lysosomal aspartyl proteases, cathepsin D (CTSD) and cathepsin E (CTSE). CTSD has been implicated in multiple neurodegenerative diseases, including a fatal congenital form of neuronal ceroid lipofuscinosis[26]. Consistent with their known pH optima, both CTSD and CTSE digested α-syn, TDP-43 and tau but only at the most acidic pH of 3.4.

Although most of the enzymes tested are endopeptidases (i.e., can cleave peptide bonds within a protein substrate), cathepsins A, C, H, and X are primarily exopeptidases, hydrolyzing mainly N- or C-terminal peptide bonds and exhibiting only limited endopeptidase activity. Within the context of our assay, we noted that cathepsins A, H, and X (exopeptidases) as well as cathepsin O (CTSO) did not demonstrate any detectable activity against α-syn, TDP-43 or tau. To confirm the endopeptidase activity of these enzymes, we performed fluorescence-based cleavage assays with fluorogenically-tagged casein, confirming that all enzymes are active and can cleave this universal substrate (**Extended Data** Fig. 1).

Multiplexed substrate profiling by mass spectrometry (MSP-MS) reveals comprehensive lysosomal protease cleavage maps of α-syn, TDP-43 and tau

To map the specific locations at which lysosomal proteases cleave α-syn, TDP-43 and tau, we took advantage of MSP-MS, a rapid, quantitative, reproducible and unbiased method of direct substrate cleavage site identification[27]. We began by designing a custom library of seven to eighteen-residue, overlapping peptides covering the entire sequences of α-syn, TDP-43 and tau (**Extended Data Table 1**). Each peptide was designed to generate unique fragments for tandem mass spectrometry (MS/MS) detection. The library was then incubated with individual proteases at one or more pH setpoints within their optimal range and cleavage fragments assessed at two time points (**Extended Data Table 2**). Cathepsins G, C and H were excluded due to reportedly low neuronal expression and because the latter two require a pre-activation step that is technically incompatible with MSP-MS[27,28]. In total, the peptide library was tested against twelve recombinant proteases which generated 920 total cleavages across the three proteins. Because the peptides within the library are small, relatively little secondary structure should exist to hinder cleavage. The results of this extensive mapping campaign are shown in **Figs. 2a, 3a and 4a** as well as in **Supplementary Data 1**.

α-Syn. At 140 amino acids, α-syn is the smallest of the three proteins assessed. In total, there were 82 cleavages in α-syn by 9 of the 12 proteases tested (**Fig. 2a**). In agreement with the recombinant, full-length α-syn *in vitro* protease assay results (Fig. 1), cathepsins A, F and O did not cleave any of the peptides within the library. Remarkably, CTSX, a protease that could not cleave full-length α-syn (potentially due to secondary structural hindrance of cleavage sites), was able to digest several α-syn peptides. Protease cleavage sites were evenly distributed throughout the α-syn sequence, including within the non-amyloid component (NAC) domain which forms the core of α-syn fibrils[2]. Of the proteases tested, CTSL and CTSS displayed the most cleavage sites in the α-syn peptides with CTSB and CTSV following (**Fig. 2b**). Of the proteases examined, cathepsins B, D and L have previously been shown to cleave α-syn[29,30]. Our results confirm cathepsins B, D and L as α-syn proteases and substantially supplement the known lysosomal proteases that can process α-syn to include six additional proteases: cathepsins E, K, S, V, X and AEP.

TDP-43. TDP-43 contains 414 amino acids. Each of the 12 proteases tested was able to cleave TDP-43 peptides in multiple locations, resulting in 553 total cleavages (**Fig. 3a**). Cathepsin S, L, V and E had the greatest number of proteolytic sites (**Fig. 3b**). Full-length, recombinant TDP-43 in solution likely retains

some secondary structure[31]. As such, it was notable that cathepsins A, F, O, X and AEP, all of which did not appear to degrade full-length TDP-43 in Fig. 1, were able to cleave a subset of the linear TDP-43 peptides under these MSP-MS conditions. Prior experimental data has demonstrated cleavage sites within TDP-43 for CTSL (positions 32, 341) and CTSS (position 341)[19]. Our results validated these sites and provided many more additional CTSL and CTSS cleavage sites. Interestingly, protease processing sites were not uniformly distributed across TDP-43. Notably, one region replete with protease sites, amino acids 306–378, forms the densely packed common core of TDP-43 fibrils[32]. In contrast, other regions within TDP-43 exhibited a relative paucity of cleavage sites, including within the carboxy (C)-terminus. Regions with a relative paucity of protease processing sites, such as the C-terminal glycine rich domain (amino acids 351 to 414) were significant, however, for AEP sites. Interestingly, AEP has previously been shown to exhibit relatively distinctive, non-overlapping cleavage sites in other proteins as well[33,34].

Tau. The longest adult isoform of tau (2N4R) contains 441 amino acids. Other than previously reported cleavage by CTSD and CTSS, little is known regarding lysosomal proteases that can digest tau[35,36]. All twelve of the proteases tested cleaved tau peptides for a total of 285 cleavages (Fig. 4A). Similar to TDP-43, cathepsins A, O and X, which did not cleave full-length recombinant tau, exhibited a few sparse cleavages of tau peptides. CTSS and CTSL displayed the highest numbers of tau cleavages with cathepsins B, V, E and AEP following (**Fig. 4b**). Also like TDP-43, protease cleavage sites within tau clustered together. Portions of tau, encompassing the proline rich P1 and P2 domains, contained relative few proteolytic sites, with residues 131 to 170 demonstrating no cleavages at all in this assay. The region from residues 306 to 355 in tau makes up the structural core of paired helical and straight filaments found in AD[4]. Again, in alignment with TDP-43, this region was relatively rich in cathepsin cleavage sites.

Proteases exhibit distinctive abilities to process α-syn, TDP-43 and tau

Cathepsins exhibit age-dependent, brain-region and cell-type specific differences in expression and activity[28,37,38], therefore we performed comparative analysis of the ability of proteases to degrade α-syn, TDP-43 and tau. Among the proteases tested, we found that two cysteine cathepsins, CTSL and CTSS, demonstrated the greatest number of cleavage sites in α-syn, TDP-43 and tau (Fig. 5A). Cathepsins B, E, V and AEP also exhibited significant activity with numerous sites of cleavage identified. Between the aspartyl proteases, CTSE exhibited significantly more cleavage sites than CTSD. Notably, cathepsins A, F and O demonstrated relatively few sites of cleavage within tau and TDP-43 and no cleavage sites within α-syn.

α-Syn, TDP-43 and tau are all intrinsically disordered proteins[39], yet we wondered if any of the proteases exhibited preferences for one substrate over the others. This prospect was of significant interest as the preferential behavior of certain lysosomal proteases could contribute to the selective neuronal vulnerability observed in different pathologies (e.g., PD versus ALS). To perform this comparative analysis in proteins of different sizes, we normalized the number of cathepsin cleavages for each substrate relative to their amino acid number. We then performed clustering analyses of the proteases

based on their ability to cleave α-syn, TDP-43 and tau[40]. This hierarchical clustering revealed up to six distinct clusters of cathepsins based on their relative preferences towards the three substrates (Fig. 5b). The first two clusters consisted of cathepsins A, F and O, which could not cleave α-syn. Cathepsins X, E and V formed a cluster that exhibit a relatively even distribution of cleavages within α-syn, TDP-43 and tau. A cluster including cathepsins L, B and S had a slight preference for α-syn cleavage while in contrast, CTSK and AEP favored TDP-43 and tau over α-syn. Lastly, cathepsins A and D rather heavily favored TDP-43 and CTSF relatively preferred tau.

Because little is known about redundant versus non-redundant activity among these proteases, we also searched for similarities among cathepsins within the cleavage site profile of each substrate. To this end, we performed pairwise correlational analyses. This analysis revealed several interesting correlations (Fig. 5c-e). Across the three substrates, the cleavage profiles of cathepsins E, L and V were positively correlated with one another. Cathepsin X was also positively correlated with CTSA and CTSO for TDP-43 and tau. Somewhat surprisingly, CTSB showed either minimal or a negative correlation with most other cathepsins in α-syn and tau. Similarly, AEP also demonstrated a negative correlation pattern with most other proteases. Overall, this analysis suggests that certain cathepsins could potentially work redundantly to degrade α-syn, TDP-43 or tau. In contrast, other cathepsins, such as CTSB and AEP, cleave relatively unique regions in these proteins. Thus, if their activity were lost, these analyses suggest that proteolytic cleavage may be affected in manner that could not be easily compensated by other proteases.

A subset of neurodegenerative disease-associated mutations is predicted to disrupt lysosomal protease cleavage

While individual proteases have preferred amino acid recognition sequences, certain positions within their recognition sequence tend to exert an outsized importance. Protease cleavage sequences are expressed based on the positions relative to the cleavage site, with (from left to right) P4 to P1 found on the amino (N)-terminal side of the cleavage site and P1’ to P4’ found on the C-terminal side. When expressed in this fashion, many of the cysteine proteases (e.g., cathepsins L and V) prefer hydrophobic residues at P2[41] while AEP prefers asparagine at P1[42]. Serine proteases rely heavily upon the P1 position for recognition[43]. Aspartyl proteases like CTSD strongly prefer hydrophobic residues at P1 and P1’[14]. Having generated cathepsin cleavage maps for α-syn, TDP-43 and tau, we overlaid the known disease-associated variants onto the maps (**Fig. 2a, 3a and 4a**). In several cases, disease-associated mutations appeared in close proximity to protease cleavage sites and were predicted to alter amino acids important for protease recognition[44]. For example, the TDP-43 A315T mutation is positioned at P1’ for a CTSE cleavage site and is predicted to make cleavage less favorable (**Fig. 3a**). Similarly, the N279K mutation on tau removes a critical asparagine from a P1 cleavage site position for AEP (**Fig. 4a**). We thus hypothesized that neurodegenerative disease-associated mutations could directly alter the efficiency of proteolytic cleavage by lysosomal proteases.

Since it was not feasible to test every mutation against every protease, we utilized two approaches to identify a subset of mutations mostly like to abrogate cathepsin cleavage. First, we performed *in silico* comparisons of WT versus mutant protein sequences in the PROSPER database (https://prosper.erc.monash.edu.au/)[8]. This analysis generated potential cleavage sites as well as a set of rank order predictions of mutations that would be likely to abrogate proteolytic cleavage (**Extended Data Fig. 2** and **Extended Data Table 3**). Second, we used the MSP-MS data to assess whether specific amino acids are preferentially recognized by certain proteases, specifically within the sequences of α-syn, TDP-43 and tau. To do so, we input the MSP-MS cleavage sites into the iceLogo algorithm (iomics.ugent.be/icelogoserver), which uses probability theory to identify conserved patterns in proteins[45]. The iceLogo output allows visualization of favorable and unfavorable amino acids at each of the protease recognition motif positions (P4 to P4') (Figs. 5f-q). Results of this analysis suggested that the amino acid preferences in α-syn, TDP-43 and tau for many of the proteases (e.g., cathepsins E, L, S and AEP) were consistent with established iceLogo maps[19,27] (Fig. 5i, l, n and **q**). However, certain proteases, in particular CTSB, showed sequence preferences within α-syn, TDP-43 and tau that diverged from current notations[19] (Fig. 5g).

From the PROSPER and iceLogo analyses, we curated a set of eleven potentially "damaging" disease mutations to test (three in α-syn, four in TDP-43, and three in tau) to test for altered proteolytic cleavage activity as well as non-damaging "control" mutations with lower probability of altering protease cleavage activity (**Extended Data Table 3**). We then generated WT and mutant peptides encompassing these regions (**Extended Data Table 4**). Each peptide was labelled with a coumarin-based (MCA) fluorophore and a fluorescence quencher on the N- and C-termini, respectively. The peptide substrates were thus quenched at baseline and only revealed fluorescence upon cleavage by a protease. We incubated the peptides with individual proteases and monitored the fluorescence generated over time. Given the nature of the assay, we limited our testing to the primary endopeptidases, cathepsins B, D, E, F, K, L, S, and V as well as AEP.

Using this activity assay, the maximal velocities (Vmax) at which each enzyme cleaved the WT and mutant peptides were calculated (**Extended Data Fig. 3**). Using these values, the rates of WT versus mutant cleavage were compared and ratios (mutant over WT) were generated. In this type of *in vitro* protease approach, every cathepsin would not be expected to cleave every peptide. Indeed, we found that each WT peptide was cleaved by between three and seven of the nine proteases tested (**Fig. 2c, 3c, 4c**, and **Extended Data Fig. 3**). Mutant peptides were either cleaved faster, similar, slower or not at all compared to wild-type peptides. Examples for α-syn, TDP-43 and tau are described below and shown in **Figs. 2c, 3c, and 4c**.

To better quantify the overall impact of these changes, we assigned "damage points" for each instance in which a mutant decreased a protease's Vmax by 0–25% (1 point), 25%-75% (2 points) or greater than 75% (3 points). In contrast, mutants that increased the rate of cleavage were given a score of -1 point. By summing the damage points, we were able to calculate a "damage score" for each mutation (**Fig. 2c, 3c,**

and 4c). Fascinatingly, each of these predicted disruptive mutations demonstrated positive scores of 3 to 6 while the controls exhibited neutral or negative scores of -3 to 2.

For α-syn peptides, the cysteine proteases CTSB, CTSL, CTSK and CTSV cleaved the mutant α-syn G51D peptide less efficiently than the WT peptide (**Fig. 2c-e**). The A53T mutation, on the other hand, diminished cleavage for CTSK, CTSL, and CTSE (**Fig. 2c, f, g**). In regard to controls, for the α-syn A30P and E46K mutant peptides, every protease capable of cleaving these sequences did so at a rate similar or faster than WT, with the exception of CTSB's activity on the A30P peptide (**Fig. 2c, h**, and **i**).

Disease-associated mutations in TDP-43 also affected proteolytic cleavage. The TDP-43 A315T mutation disrupts the important P1' position for CTSE and likely as a consequence, CTSE cleavage of the mutant A315T peptide was substantively diminished (**Fig. 3c**). TDP-43 A315T is also in proximity of other protease cleavage sites (cathepsin A, B, L, S and V) and in the *in vitro* protease assays, decreased CTSS and CTSV peptide cleavage (**Fig. 3c**). When tested *in vitro*, TDP-43 A321G substantially diminished CTSE's Vmax while more modestly diminishing CTSB, CTSD, and CTSV activity as well (**Fig. 3c and 3e**). TDP-43 Q331K is in the midst of many cleavage sites and when tested, completely abrogated CTSL and CTSV cleavage compared to WT (**Fig. 3f and 3g**) while also decreasing CTSV cleavage significantly. Similarly, the TDP-43 M337V mutation has the potential to impact many proteases, and indeed diminished the activity of cathepsins K, S and V (**Fig. 3c and 3h**). The TDP-43 G298S mutation occurs within an area of TDP-43 relatively devoid of observed protease cleavage sites. This mutation exerted generally balanced effects on protease cleavage, as CTSB and CTSK had decreased cleavage while CTSS and CTSV had an increased rate of cleavage (**Fig. 3c and 3i**).

Finally, tau mutations that were predicted to interfere with protease cleavage also dramatically impaired protease activity against mutant but not control peptides. Cathepsins F and K as well as AEP exhibited dramatically slower cleavage of the K257T mutant peptide (**Fig. 4c and 4e).** The tau N279K mutation, by removing an asparagine at an AEP P1 site, fully extinguished degradation by AEP (**Fig. 4f**) as well as Cathepsin S (**Fig. 4e**), although CTSK cleavage was increased (**Fig. 4c**). Interestingly, the S305N mutation dramatically impacted CTSE (**Fig. 4g**), CTSL (**Fig. 4h**) and CTSB (**Fig. 4c**) activity. Our negative control mutation, P301S, for which MSP-MS found limited local protease cleavage sites, demonstrated faster cleavage by CTSB that was balanced by slower cleavage of CTSV compared to WT. This is of interest as P301S and P301L are well-known mutations that rapidly promote tau aggregate formation[46] .

Disease mutations alter α-syn, TDP-43 and tau protein half-life in a neuronal model

Because the *in vitro* assays tested only a single protease at a time, we moved to a neuronal model to validate which, if any, mutations impacted lysosomal proteolysis of α-syn, TDP-43 and tau in a more physiological setting. To do so, we used SH-SY5Y cells, a validated neuroblastoma cell line that can be terminally differentiated into a neuron-like state with neurite growth and expression of neuronal markers[47,48]. We generated stable cell lines in which FLAG-tagged, full-length WT or mutant α-syn, TDP-

43 or tau could be inducibly expressed. We chose an inducible system to maintain normal neuronal physiology as much as possible, thereby limiting the observed negative effects of chronic protein overexpression on autophagy as well as mitigating the impact of these known disease mutations on neuronal health[48,49].

Following neuronal differentiation, we treated SH-SY5Y cells with doxycycline to induce α-syn, TDP-43 or tau expression for one day, then removed doxycycline and measured the rate of protein clearance. To isolate lysosomal activity, we inhibited the proteasome with MG132 and confirmed the compensatory increase in autophagic flux that has been previously seen with this inhibitor[50] (**Extended Data Fig. 4**). We then compared the clearance of WT α-syn, TDP-43 and tau to their mutants. For α-syn, the G51D and A53T mutations significantly increased protein half-life while the A30P and E46K mutations were similar to WT (Fig. 6A- B). Strikingly, these results were consistent with the predictions of the "damage scores" (**Fig. 2c, 3c, and 4c**). Similarly, TDP-43 A315T, A321G, Q331K and M337V mutations as well as tau K257T, N279K and S305N mutations all significantly prolonged the half-life of their substrates (Fig. 6c-f). Contrastingly, TDP-43 G298S and tau P301S mutations, which had low damage scores from our *in vitro* protease assays, did not extend protein half-life compared to WT. Overall, the damage scores calculated from the *in vitro* fluorogenic protease assays and the results from these half-life studies were consistently aligned. While the increase in steady-state levels of mutant proteins were subtle when measured over days, the impact of such alterations in protein half-life would be highly significant when compounded over decades and in the context of age-associated declines in lysosome function[51–53].

Discussion

We have demonstrated that single amino acid changes in client proteins can affect lysosomal proteolysis, potentially contributing to neurodegenerative disease. We first catalogued the lysosomal proteases capable of degrading α-syn, TDP43 and tau at various pH values, thereby generating a resource for future studies of lysosomal protease biology. We have also demonstrated that certain disease mutations dramatically alter proteolytic efficiency and extend protein half-life.

As the engines of autophagy, lysosomes have been increasingly recognized as key players in neurodegeneration, with mutations in lysosome-resident proteins including *GBA, PGRN* and *CTSD* implicated in PD, FTD and AD, respectively[54]. Deficient lysosomal function has also been identified as an early pathogenic event in animal models of neurodegeneration[55]. Our results further highlight lysosomal proteases as significant contributors to PD, FTD, ALS and AD pathobiology that could be differentially targeted for disease-specific therapeutic purposes. They also give rise to a novel mechanism by which both autosomal dominant mutations and coding polymorphisms in α-syn, TDP43 and tau could contribute to disease pathogenesis by interfering with lysosomal clearance, which over decades would disturb proteostasis (see model Fig. 6g). This idea is akin to how *SNCA* triplications predispose to Parkinson's disease by increasing steady state levels of a neurodegenerative disease protein[56]. Our finding that disease-associated mutations are able to decrease lysosomal proteolysis and lengthen the

half-life of these aggregating proteins therefore introduces a novel paradigm for neurodegenerative disease pathogenesis.

The implications of these findings are multifold. First, they provide a molecular basis by which disease mutations that do not promote self-association can potentially extend protein half-life, leading to accumulation and subsequent aggregate formation. Second, the specificity of protease-substrate relationships suggests a possible role for the lysosome in selective neuronal vulnerability, as protease types and levels differ between neuronal sub-types and brain regions[28] and could thereby predispose to α-syn, TDP43 or tau accumulation. Third, as cleavage site-rich regions of TDP-43 and tau overlapped with what have recently been demonstrated to form the core of TDP-43 fibrils and tau filaments, our findings suggest that failed lysosomal proteolysis may precede and contribute to inclusion formation. Finally, this approach has uncovered distinctive relationships between certain proteases and α-syn, TDP-43 or tau that could elucidate other genetic and sporadic forms of neurodegenerative disease. For example, the preference of CTSD for TDP-43 (Fig. 5b) takes on additional meaning when connected to the knowledge that CTSD maturation is promoted by the FTD-associated protein progranulin[57–59] and progranulin haploinsufficiency leads to FTD with TDP-43 inclusions[60,61]. Similarly, the proteases that cleaved TDP-43 and tau in our study were more similar than those that cleaved α-syn. This may explain why TDP-43 and tau pathology co-exist in certain cortical regions whereas α-syn pathology, at least early in disease, tends to be more regionally distinct[62].

Although our efforts were designed to be comprehensive, there are limitations to our study. While the MSP-MS and *in vitro* fluorescent protease assays are complimentary (balancing sensitivity and specificity), on occasion we observed slight discrepancies. For example, CTSB cleaved the TDP-43 fluorescent peptides spanning the 294–302 region, which we did not see by MSP-MS. This is likely because the MSP-MS results were limited to two collection time-points. As a result, a small proportion of cleavage sites for certain proteases may have not been observed or represented in our maps. Additionally, recognizing that lysosomes contain multiple proteases, it should be noted that this study was restricted to individual protease-substrate interactions as well as cell-based experiments limited to relatively short time intervals. Future work should thus include assessment of the effects of complex protease mixtures on these substrates. The contributions of post-translational modifications, which can impact both lysosomal targeting and proteolytic processing[63], also require further exploration. Evaluating the role of individual or groups of lysosomal proteases on α-syn, TDP-43, and tau degradation longitudinally within organoid or animal models represents longer term goals. Developing and testing the efficacy of lysosomal targeted protease agonists as therapies in human neurodegenerative diseases represent a future translational goal.

Conclusion

This study comprehensively identifies the arsenal of lysosomal proteases that can degrade full-length α-syn, TDP-43 and tau. These proteases have highly specific cleavage patterns and optimal pH settings.

This study provides compelling experimental evidence that certain pathogenic mutations in all three of these proteins that are primarily thought to affect protein conformation can disrupt cathepsin cleavage activity and prolong the half-life in neuronal cell models. These findings are highly relevant in understanding the consequences of understanding α-syn, TDP-43 and tau protein homeostasis and could be leveraged to promote neuronal health through augmentation of the activity of specific cathepsins.

Declarations

Data Availability:

The authors declare that all data supporting the findings of this study are available within the paper and its supplementary information files.

ACKNOWLEDGEMENTS:

We would like to thank the members of the Kao and Jacobson labs at UCSF for reviewing this manuscript.

FUNDING:

Richard Olney Clinician Scientist Development Award from the American Academy of Neurology, American Brain Foundation, and ALS Association (to PJS); National Institute of Health (NIH) R01NS095257, R01 AG057342, R01AG059052, The Paul G. Allen Family Foundation, The Rainwater Tau Consortium, Parkinson's Spectrum Disease Center, Glenn Foundation for Aging Research (to AWK); the Larry Berkelhammer Foundation (to MPJ); NIA R01AG059052, NIH RF1AG068125 (to CSC).

AUTHOR CONTRIBUTIONS:

P.J.S., S.A., M.P.J. and A.W.K. conceived the project. P.J.S., S.A., G.M.K., C.S.C. and A.W.K. designed all experiments. P.J.S. performed *in vitro* silver stain experiments as well as the non-cleaving enzyme activity assays. P.J.S. and S.A. performed all protease activity assay experiments. G.M.K. performed the mass spectrometry experiments. G.M.K., S.A. and P.J.S. performed the mass spectrometry analysis. E.L.G. and A.S.P. performed the in vitro protease activity assay modeling and analysis. S.A. performed the hierarchical and correlational analyses. P.J.S. and S.A. performed the cell-based studies. P.J.S., S.A., and A.W.K. composed the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS:

A.W.K. and M.P.J. are Scientific Advisory Board members, and M.P.J. is a founder of Nine Square Therapeutics. C.S.C. is a founder, and G.M.K. is an employee of Alaunus Biosciences. Authors declare that they have no conflicts of interest or competing interests.

ADDITIONAL INFORMATION:

Supplementary Information is available for this paper

MATERIALS & CORRESPONDENCE

Correspondence and requests for materials should be addressed to Dr. Aimee Kao (aimee.kao@ucsf.edu).

References

1. Dugger, B. N. & Dickson, D. W. Pathology of Neurodegenerative Diseases. Cold Spring Harb Perspect Biol 9, a028035 (2017).
2. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat Commun* **9**, 3609 (2018).
3. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat Struct Mol Biol* **26**, 619–627 (2019).
4. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
5. Mathieu, C., Pappu, R. V. & Taylor, J. P. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* **370**, 56–60 (2020).
6. Hipp, M. S., Kasturi, P. & Hartl, F. U. The proteostasis network and its decline in ageing. *Nat Rev Mol Cell Biol* **20**, 421–435 (2019).
7. Rawlings, N. D. Using the MEROPS Database for Investigation of Lysosomal Peptidases, Their Inhibitors, and Substrates. *Methods Mol Biol* **1594**, 213–226 (2017).
8. Song, J. *et al.* PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One* **7**, e50300 (2012).
9. Biniossek, M. L., Nägler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J Proteome Res* **10**, 5363–5373 (2011).
10. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat Rev Neurosci* **14**, 38–48 (2013).
11. Gao, J., Wang, L., Huntley, M. L., Perry, G. & Wang, X. Pathomechanisms of TDP-43 in neurodegeneration. *J Neurochem* (2018).
12. Gao, Y.-L. *et al.* Tau in neurodegenerative disease. *Ann Transl Med* **6**, 175 (2018).
13. Burré, J., Sharma, M. & Südhof, T. C. Cell Biology and Pathophysiology of α-Synuclein. *Cold Spring Harb Perspect Med* **8**, a024091 (2018).
14. Blokhuis, A. M., Groen, E. J. N., Koppers, M., van den Berg, L. H. & Pasterkamp, R. J. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol* **125**, 777–794 (2013).

15. Strang, K. H., Golde, T. E. & Giasson, B. I. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. *Lab Invest* **99**, 912–928 (2019).
16. Leeman, D. S. *et al.* Lysosome activation clears aggregates and enhances quiescent neural stem cell activation during aging. *Science* **359**, 1277–1283 (2018).
17. Perera, R. M. & Zoncu, R. The Lysosome as a Regulatory Hub. *Annu Rev Cell Dev Biol* **32**, 223–253 (2016).
18. Metcalf, D. J., García-Arencibia, M., Hochfeld, W. E. & Rubinsztein, D. C. Autophagy and misfolded proteins in neurodegeneration. *Exp Neurol* **238**, 22–28 (2012).
19. Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat Med* **21**, 1406–1415 (2015).
20. Banay-Schwartz, M. *et al.* The pH dependence of breakdown of various purified brain proteins by cathepsin D preparations. *Neurochem Int* **7**, 607–614 (1985).
21. Stefanis, L. *et al.* How is alpha-synuclein cleared from the cell? *J Neurochem* **150**, 577–590 (2019).
22. Huang, C.-C. *et al.* Metabolism and mis-metabolism of the neuropathological signature protein TDP-43. *J Cell Sci* **127**, 3024–3038 (2014).
23. Lee, M. J., Lee, J. H. & Rubinsztein, D. C. Tau degradation: the ubiquitin-proteasome system versus the autophagy-lysosome system. *Prog Neurobiol* **105**, 49–59 (2013).
24. Turk, V. *et al.* Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta* **1824**, 68–88 (2012).
25. Wilder, C. L., Park, K.-Y., Keegan, P. M. & Platt, M. O. Manipulating substrate and pH in zymogr aphy protocols selectively distinguishes cathepsins K, L, S, and V activity in cells and tissues. *Arch Biochem Biophys* **516**, 52–57 (2011).
26. Siintola, E. *et al.* Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain* **129**, 1438–1445 (2006).
27. O'Donoghue, A. J. *et al.* Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods* **9**, 1095–1100 (2012).
28. Hsu, A., Podvin, S. & Hook, V. Lysosomal Cathepsin Protease Gene Expression Profiles in the Human Brain During Normal Development. *J Mol Neurosci* **65**, 420–431 (2018).
29. Sevlever, D., Jiang, P. & Yen, S.-H. C. Cathepsin D is the main lysosomal enzyme involved in the degradation of alpha-synuclein and generation of its carboxy-terminally truncated species. *Biochemistry* **47**, 9678–9687 (2008).
30. McGlinchey, R. P. & Lee, J. C. Cysteine cathepsins are essential in lysosomal degradation of α-synuclein. *Proc Natl Acad Sci U S A* **112**, 9322–9327 (2015).
31. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24**, 1537–1549 (2016).
32. Li, Q., Babinchak, W. M. & Surewicz, W. K. Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat Commun* **12**, 1620 (2021).

33. Zhang, Z., Xie, M. & Ye, K. Asparagine endopeptidase is an innovative therapeutic target for neurodegenerative diseases. *Expert Opin Ther Targets* **20**, 1237–1245 (2016).
34. Mohan, S. *et al.* Processing of progranulin into granulins involves multiple lysosomal proteases and is affected in frontotemporal lobar degeneration. *Mol Neurodegener* **16**, 51 (2021).
35. Kenessey, A., Nacharaju, P., Ko, L. W. & Yen, S. H. Degradation of tau by lysosomal enzyme cathepsin D: implication for Alzheimer neurofibrillary degeneration. *J Neurochem* **69**, 2026–2038 (1997).
36. Nübling, G. *et al.* Cathepsin S increases tau oligomer formation through limited cleavage, but only IL-6, not cathespin S serum levels correlate with disease severity in the neurodegenerative tauopathy progressive supranuclear palsy. *Exp Brain Res* **235**, 2407–2412 (2017).
37. Nakanishi, H. *et al.* Increased expression of cathepsins E and D in neurons of the aged rat brain and their colocalization with lipofuscin and carboxy-terminal fragments of Alzheimer amyloid precursor protein. *J Neurochem* **68**, 739–749 (1997).
38. Ni, J. *et al.* The Critical Role of Proteolytic Relay through Cathepsins B and E in the Phenotypic Change of Microglia/Macrophage. *J Neurosci* **35**, 12488–12501 (2015).
39. Zbinden, A., Pérez-Berlanga, M., De Rossi, P. & Polymenidou, M. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Developmental Cell* **55**, 45–68 (2020).
40. Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31**, 651–666 (2010).
41. Puzer, L. *et al.* Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L. *Arch Biochem Biophys* **430**, 274–283 (2004).
42. Schwarz, G. *et al.* Characterization of legumain. *Biol Chem* **383**, 1813–1816 (2002).
43. Hedstrom, L. Serine protease mechanism and specificity. *Chem Rev* **102**, 4501–4524 (2002).
44. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067 (2018).
45. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786–787 (2009).
46. Strang, K. H. *et al.* Distinct differences in prion-like seeding and aggregation between Tau protein variants provide mechanistic insights into tauopathies. *J Biol Chem* **293**, 2408–2421 (2018).
47. Xicoy, H., Wieringa, B. & Martens, G. J. M. The SH-SY5Y cell line in Parkinson's disease research: a systematic review. *Mol Neurodegener* **12**, 10 (2017).
48. Nascimento, A. C. *et al.* α-Synuclein Overexpression Induces Lysosomal Dysfunction and Autophagy Impairment in Human Neuroblastoma SH-SY5Y. *Neurochem Res* **45**, 2749–2761 (2020).
49. Austin, J. A. *et al.* Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc Natl Acad Sci U S A* **111**, 4309–4314 (2014).
50. Li, C. *et al.* Proteasome Inhibition Activates Autophagy-Lysosome Pathway Associated With TFEB Dephosphorylation and Nuclear Translocation. *Front Cell Dev Biol* **7**, 170 (2019).

51. Hughes, A. L. & Gottschling, D. E. An early age increase in vacuolar pH limits mitochondrial function and lifespan in yeast. *Nature* **492**, 261–265 (2012).
52. Folick, A. *et al.* Aging. Lysosomal signaling molecules regulate longevity in Caenorhabditis elegans. *Science* **347**, 83–86 (2015).
53. Sun, Y. *et al.* Lysosome activity is modulated by multiple longevity pathways and is important for lifespan extension in C. elegans. *Elife* **9**, e55745 (2020).
54. Root, J., Merino, P., Nuckols, A., Johnson, M. & Kukar, T. Lysosome dysfunction as a cause of neurodegenerative diseases: Lessons from frontotemporal dementia and amyotrophic lateral sclerosis. *Neurobiology of Disease* **154**, 105360 (2021).
55. Lee, Y. *et al.* Early lysosome defects precede neurodegeneration with amyloid-β and tau aggregation in NHE6-null rat brain. *Brain* awab467 (2021).
56. Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
57. Butler, V. J. *et al.* Progranulin Stimulates the In Vitro Maturation of Pro-Cathepsin D at Acidic pH. *J Mol Biol* **431**, 1038–1047 (2019).
58. Butler, V. J. *et al.* Age- and stress-associated C. elegans granulins impair lysosomal function and induce a compensatory HLH-30/TFEB transcriptional response. *PLoS Genet* **15**, e1008295 (2019).
59. Zhou, X. *et al.* Regulation of cathepsin D activity by the FTLD protein progranulin. *Acta Neuropathol* **134**, 151–153 (2017).
60. Baker, M. *et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442**, 916–919 (2006).
61. Cruts, M. & Van Broeckhoven, C. Loss of progranulin function in frontotemporal lobar degeneration. *Trends Genet* **24**, 186–194 (2008).
62. James, B. D. *et al.* TDP-43 stage, mixed pathologies, and clinical Alzheimer's-type dementia. *Brain* **139**, 2983–2993 (2016).
63. Alquezar, C. *et al.* TSC1 loss increases risk for tauopathy by inducing tau acetylation and preventing tau clearance via chaperone-mediated autophagy. *Sci Adv* **7**, eabg3897 (2021).

Methods

Recombinant proteases

Cathepsin E (CSTE) (R&D #1294-AS), cathepsin D (CTSD) (R&D #1014-AS), cathepsin G (CTSG) (Millipore #219873), cathepsin A (CTSA) (R&D #1049-SE), cathepsin L (CSTL) (Millipore #219402), cathepsin B (CSTB) (Millipore #219364), cathepsin K (CTSK) (Millipore #219461), (CTSS) cathepsin S (R&D #1183-CY), cathepsin V (CTSV) (R&D #1080-CY), asparagine endopeptidase (AEP) (R&D #2199-CY), cathepsin H (CTSH) (R&D #7516-CY-010), cathepsin C (CTSC) (R&D #1071-CY), cathepsin O (CTSO) (Abcam #ab267932), cathepsin F (CTSF) (Abcam #ab240858), and cathepsin X (CTSX) (R&D #934-CY).

Antibodies

The following primary antibodies were used: 1) monoclonal mouse anti-GAPDH (Abcam, #ab8245, 1:2500) and 2) primary mouse anti-flag antibody (F3040; Sigma-Aldrich) at a 1:1000 dilution.

Plasmid constructs, point mutations, and stable lines

Novel doxycycline-inducible, flag-tagged, full length WT Tau, TDP-43, and α-syn lentiviral plasmid constructs were first designed and synthesized (Epoch Life Sciences) using a puromycin resistant plasmid backbone, pTet-O-Ngn2-Puro (Addgene #52047). Using these 3 WT plasmid constructs, individual constructs containing single pathogenic point mutations were also generated. Tau: K257T, N279K, P301S, and S305N. TDP-43: G298S, A315T, A321G, Q331K, and M337V. A-syn: A30P, E46K, H50Q, G51D, and A53T. With these constructs, lentivirus was made using the psPAX2 packing (Addgene #12260) and pCMV-VSV-G envelope (Addgene #8454) plasmids. To generate stable lines, SH-SY5Y cells already containing the pLenti CMV rtTA3 Blast construct (Addgene #26429) were infected and stable lines (17 in total) were generated using Puromycin 1 mg/mL selection.

In vitro protease cleavage assays

For *in vitro* cleavage assays, 1 µg of recombinant full-length human 4N2R tau (rPeptide #T-1001-1), full-length TDP-43 (R&D #AP-190), or full-length α-syn (Abcam #ab51189) was incubated with or without 1µM of each protease. Proteases requiring pre-activation were performed, as mentioned in **Extended Data Table 2**. Depending on the pH setpoint, the following buffers were used: 100 mM sodium citrate pH 3.4, 50 mM sodium acetate pH 4.5 or 5.5 or 100 mM phosphate buffer saline (PBS) pH 7.4. 1mM EDTA and 2mM DTT were also used, and each reaction was performed over 1 hour at 37ºC. The assay was performed in a total volume of 19.5µl. Protease activity was stopped by adding 7.5µl of NuPAGE 4X LDS (Fisher Scientific #NP0007) and 3µl of 10X reducing agent (i.e., 50µM) (Fisher #NP0009). Samples were then immediately denatured for 10 minutes at 80ºC. All samples were run on precast NOVEX 4-12% Bis-Tris gels (Fisher #NP0321PK2) using MES buffer (Fisher #NP0002). The gel was then either fixed in 40% ethanol and 10% acetic acid for silver or transferred onto nitrocellulose membranes for western blotting analysis. Silver staining was performed according to manufacturer's instructions with SilverQuest silver staining kit (Thermo Fisher #LC6070).

Multiplex substrate profiling by mass spectrometry (MSP-MS)

MSP-MS was performed as previously published by O'Donoghue et al., 2012[27]. A substrate library was designed to cover the sequences for the three human proteins TDP-43, α-syn, and tau. The specific isoforms, their protein length and the UniProt accession numbers selected were as follows: TDP-43 isoform 1 (414 aa, Q13148), α-syn isoform 1 (140 aa, P37840), and tau 2N4R isoform Tau-F (441 aa, P10636-8). To design library peptides, overlapping fragments were chosen in a tiling approach, generally using a length of 18 amino acids and 5 amino acid overlap between fragments . In some cases, highly acidic regions such as 105 – 140 in α-syn, containing many Asp and Glu residues, would have yielded

peptides bearing too much negative charge; these fragments were designed with shorter overlapping fragments of length 7 – 16 amino acids. To avoid oxidation of Cys residues that can form non-natural aggregates, all Cys sites were mutated to Ala. Finally, additional basic and spacing residues were appended to the N- and C-termini of each peptide to produce peptides suitable for liquid chromatography tandem mass spectrometric (LC-MS/MS) analysis. The final substrate library contained an equimolar mixture of 77 peptide substrates with a maximum length of 24 amino acids, with the sequences provided in **Extended Data Table 1**.

MSP-MS reactions were prepared with proteases at 1-70 nM concentration with a library concentration of 500 nM for each peptide (except CTSF at 1 μM) and incubated at 37 °C. Buffers were sodium acetate buffer (50 mM) containing 5 mM DTT and 1 mM EDTA for pH 4.5 and pH 5.5, or sodium citrate (50 mM), containing 5 mM DTT and 1 mM EDTA at pH 3.5. Reactions were monitored at two time points in an end point screening format. Each aliquot taken at a given time point was immediately desalted with C18 zip tips (Millipore-Sigma), and then freeze-dried. Samples were re-suspended in 0.1% formic acid in HPLC-grade water for LC-MS/MS analysis.

Reaction conditions for each protease followed manufacturer recommendations for pre-activation and then were assayed at the concentrations and pH buffer conditions mentioned in **Extended Data Table 2**. For CTSA, a matched CTSL-only sample was also prepared for this reaction for use as a negative control in cleavage identification. The final CTSA reaction contained 41 nM CTSA with a background of 8.25 nM CTSL, treated with E64 (440 nM), and the matched reaction contained only CTSL and E64.

Mass spectrometry

Peptide sequencing by LC-MS/MS was performed on an QExactive Plus mass spectrometer (Thermo), equipped with a nanoACQUITY (Waters) ultraperformance LC (UPLC) system and an EASY-Spray ion source (Thermo). Reversed-phase chromatography was carried out with an EASY-Spray PepMap C18 column (Thermo, ES800; 3μm bead size, 75μm by 150 mm). Chromatography was performed at a 600-nl/min flow rate during sample loading for 14 min, and then at a 400-nl/min flow rate for peptide separation over 90 min with a linear gradient of 2 to 35% (vol/vol) acetonitrile in 0.1% formic acid. Peptide fragmentation was performed by higher-energy collisional dissociation (HCD) on the six most intense precursor ions, with a minimum of 2,000 counts, with an isolation width of 2.0 m/z and a minimum normalized collision energy of 25. Data were analyzed using Protein Prospector software, v.6.2.1 (http://prospector.ucsf.edu/prospector/mshome.htm, UCSF) using published methods[27]. The peptide cleavage data were then output as 8-mer sequences that spanned the P4 – P4’ sites for each verified cleavage site (**Supplementary Data 1**).

Hierarchical clustering of protein level cleavages from MSP-MS

The percentage contribution of each enzyme to the total number of cleavages in each protein was calculated by dividing the number of cleavage sites for each enzyme with the sum of all enzyme cleavage sites. Since our substrates α-syn, TDP-43 and tau have different amino acid lengths, in order to

compare the cleavage patterns, we next normalized the percentage contribution of enzyme cleavages for each substrate with the mean of total percentage contribution for each enzyme across all substrates. Using these mean-normalized values for each enzyme within each substrate, we then performed an unsupervised hierarchical clustering in Python using the clustermap function within the data visualization library called “seaborn” (https://seaborn.pydata.org/index.html).

Correlation analysis of enzyme cleavage profiles

In order to compare the cleavage profile for each enzyme within each protein substrate, we computed the correlation matrix with the raw cleavage data obtained from MSP-MS and then plotted this matrix as a heatmap using the correlation function within seaborn in Python. The diagonal correlation matrix was then plotted in the form of heatmap.

In silico analysis of MSP-MS data

The resulting peptide identifications via MSP-MS were assessed for specific cleavage in the enzyme-treated sample by subtracting the results from a no-enzyme control incubation. Both endopeptidase and exopeptidase cleavage sites were readily identified. Using these sites, we generated full sequence cleavage maps of α-syn, TDP-43 and tau. We also created IceLogos diagrams of each cathepsin to identify which amino acids were favored at the P4’ through P4 positions of each enzyme’s specific cleavage motif[45]. We further validated our findings using the PROSPER protease prediction algorithm server (https://prosper.erc.monash.edu.au/)[8]. To determine which mutations might alter protease cleavage, we also conferred with the MEROPS database (https://www.ebi.ac.uk/merops/), assessing the amino acid frequency at the P4’-P4 positions, where all known cathepsins (cathepsins A, B, D, E, F, K, L, O, S, V, X and AEP) were previously found to cleave[7]. Using the results of our MEROPS and PROSPER searches as well as our experimental results, we identified regions of α-syn, TDP-43 and tau where pathogenic mutations overlapped with protease cleavage sites. Using this information, we designed fluorogenic peptides to test for our activity assays.

Protease activity assays

Fluorescence-based protease activity assays were performed in triplicates in black, flat-bottom 384-well plates (Greiner #781091, Fisher Scientific) using custom designed fluorescent α-syn, TDP-43 and tau peptide substrates (Genscript, **Extended Data Table 3**). Assays were run at 37°C for 10 hours in 50 mM sodium citrate (pH 3.4), 50 mM sodium acetate (pH 4.5 and 5.5) or 50 mM phosphate buffer saline (pH 7.4) buffers containing 2mM DTT and 1mM EDTA. The concentration of cathepsins and fluorescence substrate in all the assays was 20 nM and 20 μM, respectively. The substrates were designed in the form of a fluorophore/quencher paired lysine conjugated with 7-methoxycoumarin-4-acetic acid (MCA) as a fluorophore on the N-terminus, and a lysine conjugated to dinitrophenol (DNP) as a quencher on the C-terminus. In addition, two arginine residues were added to all the substrates in order to increase the solubility of peptides in the buffers. When the substrate is intact, MCA is non-fluorescent due to the presence of the DNP quencher. However, when the substrate is cleaved by means of protease activity,

MCA becomes unquenched and fluorescent. This MCA fluorescence, which serves a readout for substrate cleavage, was monitored as a function of time in a Tecan Infinite M Plex plate reader using excitation and emission wavelengths of 328nm and 393nm, respectively.

Raw MCA fluorescence data was normalized from 0-2000 using Microsoft Excel Version 16.49. Normalized data was then input into a kinetic model using Berkeley Madonna Version 10.2.8. to fit the enzyme kinetic data using the Rosenbrock Stiff method. Equations of the model were as follows:

d/dt (C1) = -a*C1*C1-b*C1*C

d/dt (C) = +a*C1*C1+b*C1*C

d/dt (S) = -k*C*S

d/dt (P) = k*C*S

C1 represented the proform of the cathepsin, C represented the mature cathepsin, S represented the full custom designed fluorescent α-syn, TDP-4, and tau peptide substrates, and P represented the cleaved substrate. Initial concentrations of the proform cathepsin and the substrate were 2 nM and 2000 nM, respectively to represent experimental concentrations. Parameters a & b represented cathepsin maturation. Parameter a represented the slow auto-activation mechanism by the proenzyme, while parameter b represented the faster rate of activation that occurs by the mature cathepsin. Initial estimates for parameter a were 0.05 and 0.15, and 0.5 and estimate for parameter b was 1.5. Parameter k represents substrate cleavage. Initial estimates for parameter k were 5.0e-4 and 0.0015. All parameters were constrained to be positive values. Normalized MCA fluorescence data was fit to variable P.

Cell-based protein half-life experiments

SH-SY5Y human neuroblastoma cells were obtained from ATCC (CRL-2266) and maintained 1:1 EMEM/F12 media (ATCC #30-2003/Thermo #11765062) supplemented with 10% FBS and 1% Penicillin-Streptomycin (Thermo Fisher #15140122). Stable lines of all 17 aformentioned constructs were generated by lentiviral infection followed by both blasticidin (5 µg/mL) and puromycin (1mg/mL) selection. To terminally differentiate into neuron-like cells, SH-SY5Y cells were treated as published[34]. Once differentiated, SH-SY5Y cells were treated with 1mg/mL of doxycycline to induce expression of Flag-tagged α-syn, TDP-43 or tau for 24 hours. On day 2, the media was refreshed without doxycycline and cells were harvested at 0 , 96 or 192 hrs. For inhibitor studies, differentiated SH-SY5Y cells had their media refreshed and supplemented with either 10µM PBS (control), 100nm MG132 (EMD Millipore #474791), 10µM Pepstatin A (Sigma #P5318), 10µM E64D (Tocris #4545), or 20nM Bafilomycin A1 (Tocris #1334).

Western blotting was subsequently performed as previously published[34]. 40µg of protein were separated on 4–12% SDS-PAGE (Thermo) and then transferred onto nitrocellulose membranes (Bio-Rad). Membranes were blocked at room temperature with Odyssey Blocking Buffer (LI-COR 927-40100) and

incubated at 4 °C overnight with primary antibodies and 1 hour at room temperature with appropriate fluorescent secondary antibodies (1:5000) (LI-COR). Immunoreactive bands were visualized using a LI-COR Odyssey CLx image scanner and quantified using ImageJ software.

Statistical analysis

Details of the statistical test used for each experiment is in figure legends along with n and p value. All data is represented as mean ± SD. Statistical analysis was performed using GraphPad Prism 9 (GraphPad Software, La Jolla, California USA). Western blot quantifications for **Figures 6B, D,** and **E** as well as **Extended Data Figure 4** were analyzed using 2-way, repeated measures ANOVA analyses. Pearson's correlational coefficient testing was used for the pairwise correlational analyses for Figure 5C-E. A p-value < 0.05 was considered significant. *p<0.05, **p<0.01, or ***p <0.001.

Figures

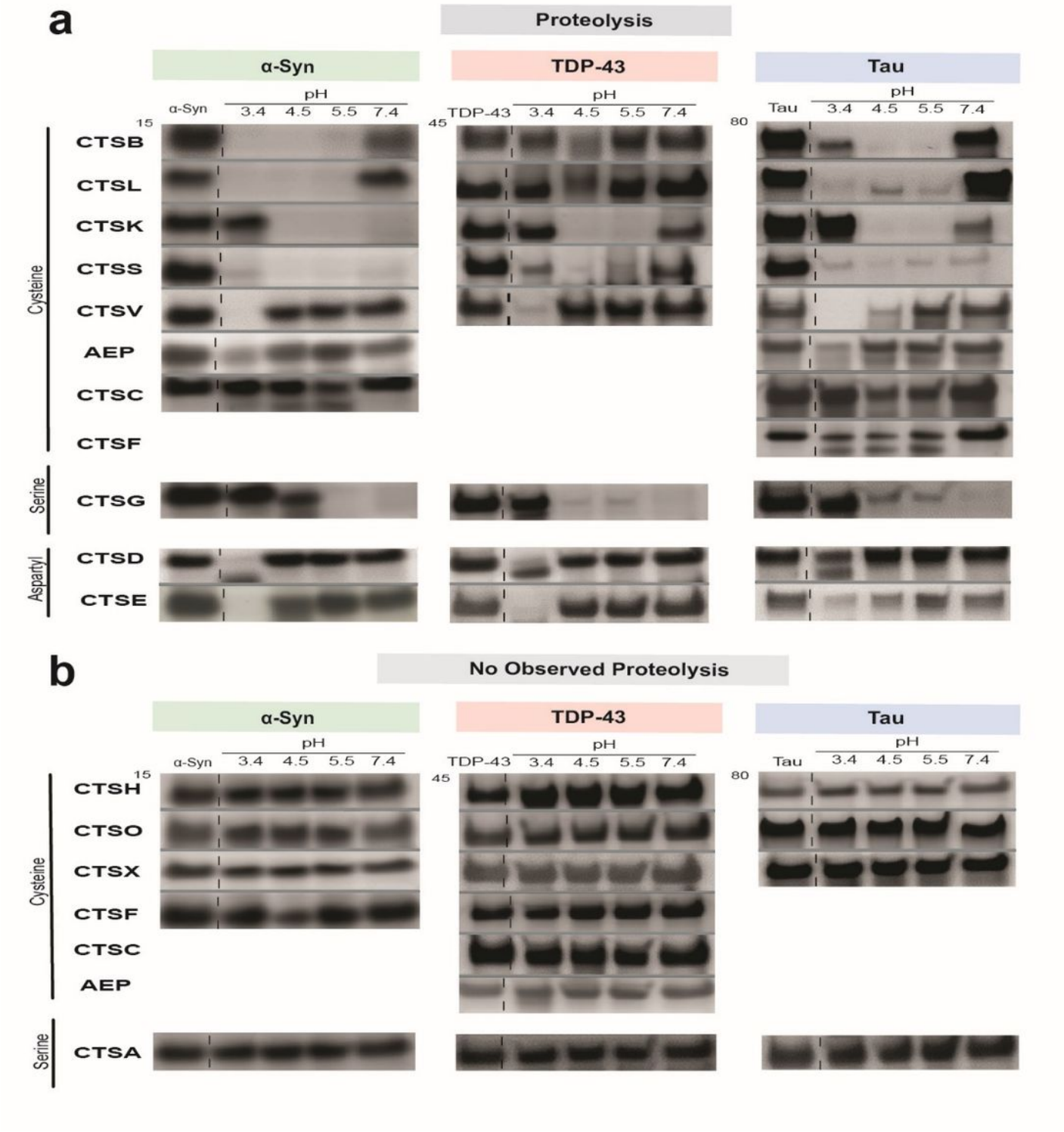

Figure 1

Lysosomal proteases digest recombinant human α-syn, TDP-43 and tau in a selective fashion. a-b, 1 μM of each enzyme was incubated with 1 μg of recombinant, human, full-length α-syn, TDP-43 or 2N4R tau at the indicated pH for 1 hour. Reactions were subjected to SDS-PAGE and gels visualized by silver stain to assess the presence (A) or absence (B) of substrate proteolysis.

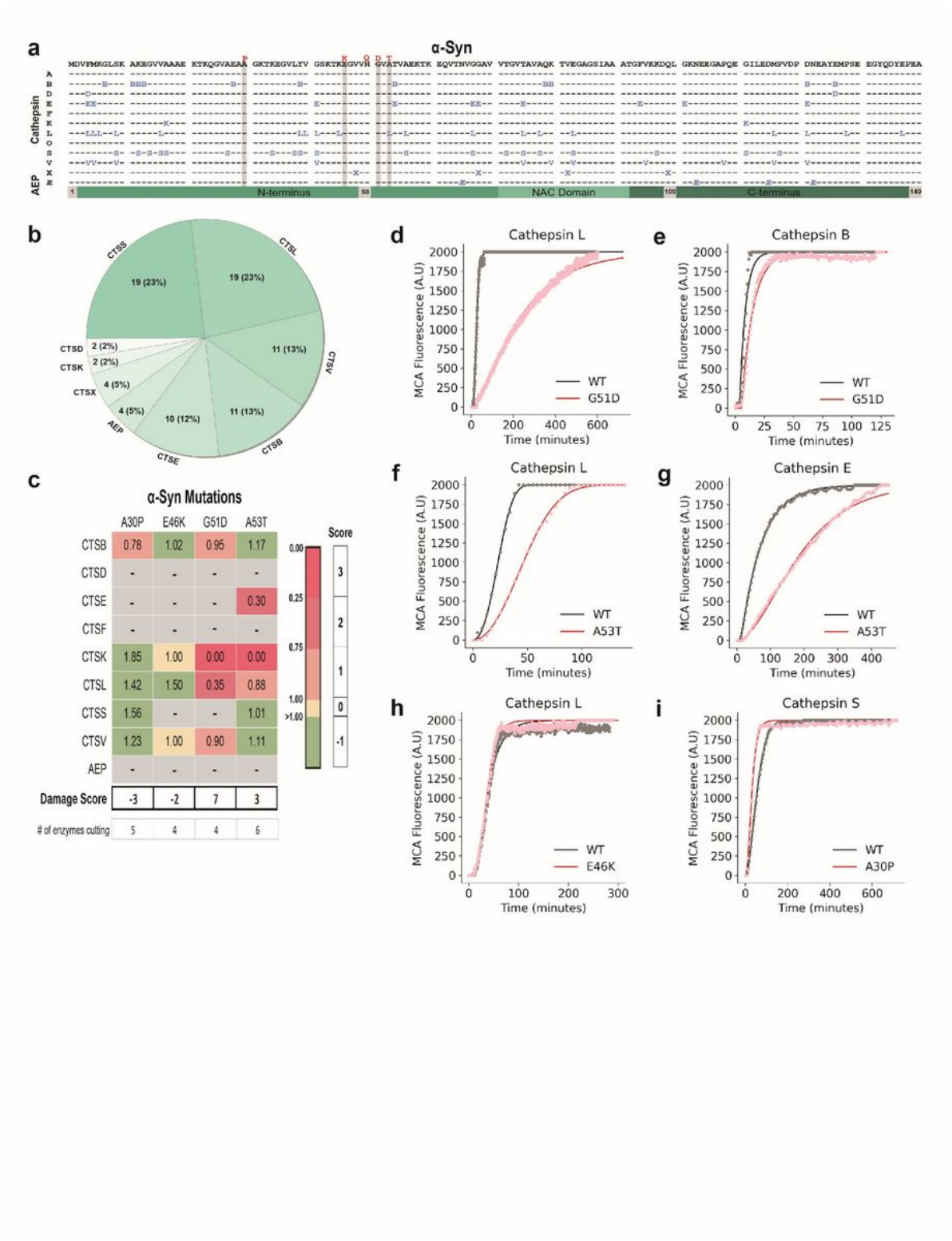

Figure 2

α-Syn proteolysis map and *in vitro* fluorescence protease assays. a, Twelve lysosomal proteases were tested for the ability to cleave α-syn peptides via the MSP-MS assay, and the results are shown here. The amino acid sequence of α-syn is in black letters at the top. Enzymes are listed to the left. If the enzyme has a cleavage site in α-syn, it is indicated with the enzyme letter (e.g., B for CTSB) positioned below and to the left of the cleavage site, in the P1 position. A total of 82 cleavages were found. Autosomal-

dominant coding mutations associated with Parkinson's Disease are noted in red above the α-syn sequence. Grey bars highlight amino acid mutations tested in *in vitro* fluorescent protease assays below. **b**, Pie chart demonstrating the number of cleavage sites within the α-syn sequence for each protease with the percentage of contributed cleavage sites in parentheses. **c**, Table of maximal velocity (Vmax) ratios (mutant/WT), comparing protease cleavage of WT versus mutant α-syn peptides. Mutations disrupting protease cleave by 0-25% (1 point), 25-75% (2 points), and > 75% (3 points) are highlighted in light pink, dark pink, and dark red, respectively. Mutations augmenting the rate cleavage (-1 point) are highlighted in light green. Mutations with similar rate of cleavage compared to WT (0 points) are highlighted in yellow. Grey boxes denote no observed cleavage for either the WT or mutant peptide. Points for each mutation were summed and a total "damage score" was given. **d-i,** Representative curves of fluorescence generated from α-syn peptide cleavage, comparing WT and mutant peptides as labeled. NAC, non-amyloid component domain; Æ, arginine endopeptidase (AEP).

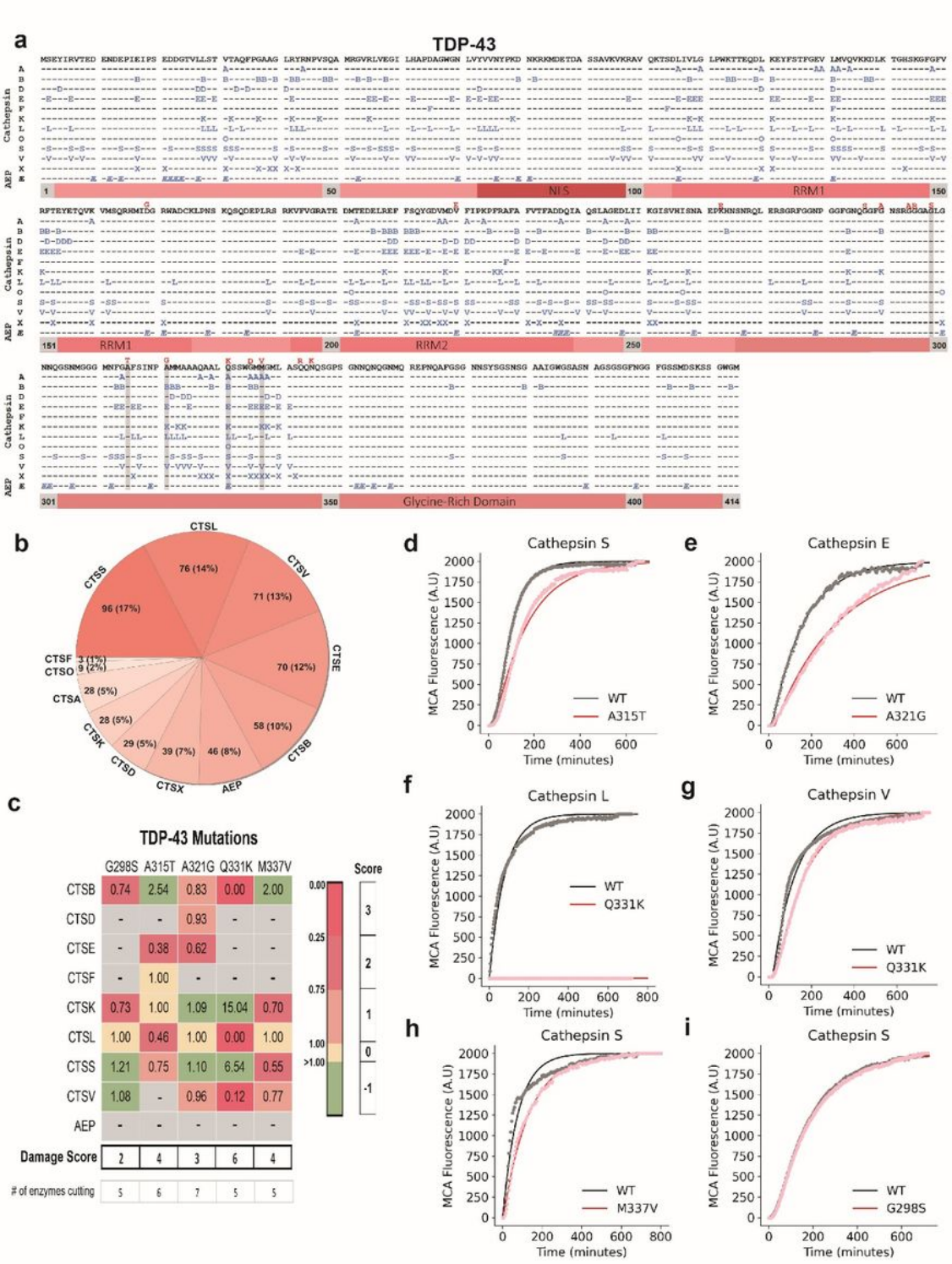

Figure 3

TDP-43 proteolysis map and *in vitro* fluorescence protease assays. a, Twelve lysosomal proteases were tested for the ability to cleave TDP-43 peptides via the MSP-MS assay, and the results are shown here. The amino acid sequence of TDP-43 is in black letters at the top. Enzymes are listed to the left. If the enzyme has a cleavage site in TDP-43, it is indicated with the enzyme letter (e.g., B for CTSB) positioned below and to the left of the cleavage site, in the P1 position. A total of 553 cleavages were found.

Autosomal-dominant coding mutations associated with amyotrophic lateral sclerosis are noted in red above the TDP-43 sequence. Grey bars highlight amino acid mutations tested in *in vitro* fluorescent protease assays below. **b**, Pie chart demonstrating the number of cleavages sites within the TDP-43 sequence for each protease with the percentage of contributed cleavage sites in parentheses. **c**, Table of maximal velocity (Vmax) ratios (mutant/WT), comparing protease cleavage of WT versus mutant TDP-43 peptides. Mutations disrupting protease cleave by 0-25% (1 point), 25-75% (2 points), and > 75% (3 points) are highlighted in light pink, dark pink, and dark red, respectively. Mutations augmenting the rate cleavage (-1 point) are highlighted in light green. Mutations with similar rate of cleavage compared to WT (0 points) are highlighted in yellow. Grey boxes denote no observed cleavage for either the WT or mutant peptide. Points for each mutation were summed and a total “damage score” was given. **d-i,** Representative curves of fluorescence generated from TDP-43 peptide cleavage, comparing WT and mutant peptides as labeled. NLS, nuclear localization sequence; RRM1 and RRM2, RNA recognition motifs 1 and 2; Æ, arginine endopeptidase (AEP).

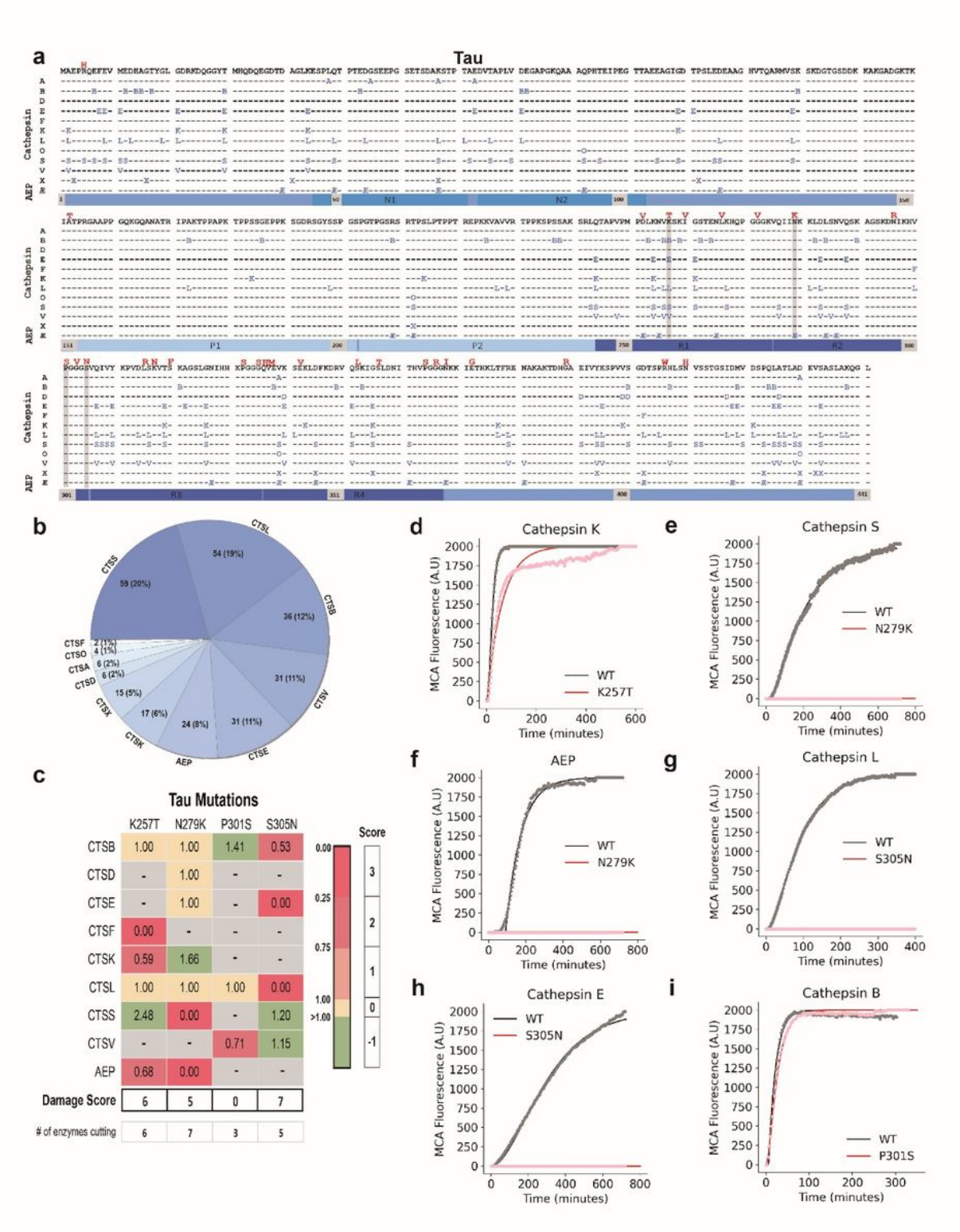

Figure 4

Tau proteolysis map and *in vitro* fluorescence protease assays. a, Twelve lysosomal proteases were tested for the ability to cleave tau peptides via the MSP-MS assay, and the results are shown here. The amino acid sequence of tau is in black letters at the top. Enzymes are listed to the left. If the enzyme has a cleavage site in tau, it is indicated with the enzyme letter (e.g., B for CTSB) positioned below and to the left of the cleavage site, in the P1 position. A total of 285 cleavages were found. Autosomal-dominant

coding mutations associated with frontotemporal dementia are noted in red above the tau sequence. Grey bars highlight amino acid mutations tested in *in vitro* fluorescent protease assays below. **b**, Pie chart demonstrating the number of cleavages sites within the tau sequence for each protease with the percentage of contributed cleavage sites in parentheses. **c**, Table of maximal velocity (Vmax) ratios (mutant/WT), comparing protease cleavage of WT versus mutant tau peptides. Mutations disrupting protease cleave by 0-25% (1 point), 25-75% (2 points), and > 75% (3 points) are highlighted in light pink, dark pink, and dark red, respectively. Mutations augmenting the rate cleavage (-1 point) are highlighted in light green. Mutations with similar rate of cleavage compared to WT (0 points) are highlighted in yellow. Grey boxes denote no observed cleavage for either the WT or mutant peptide. Points for each mutation were summed and a total “damage score” was given. **d-i,** Representative curves of fluorescence generated from tau peptide cleavage, comparing WT and mutant peptides as labeled. N1 and N2, N-terminal repeats; P1 and P2, proline-rich regions, R1-4, microtubule binding repeats 1-4; Æ, arginine endopeptidase (AEP).

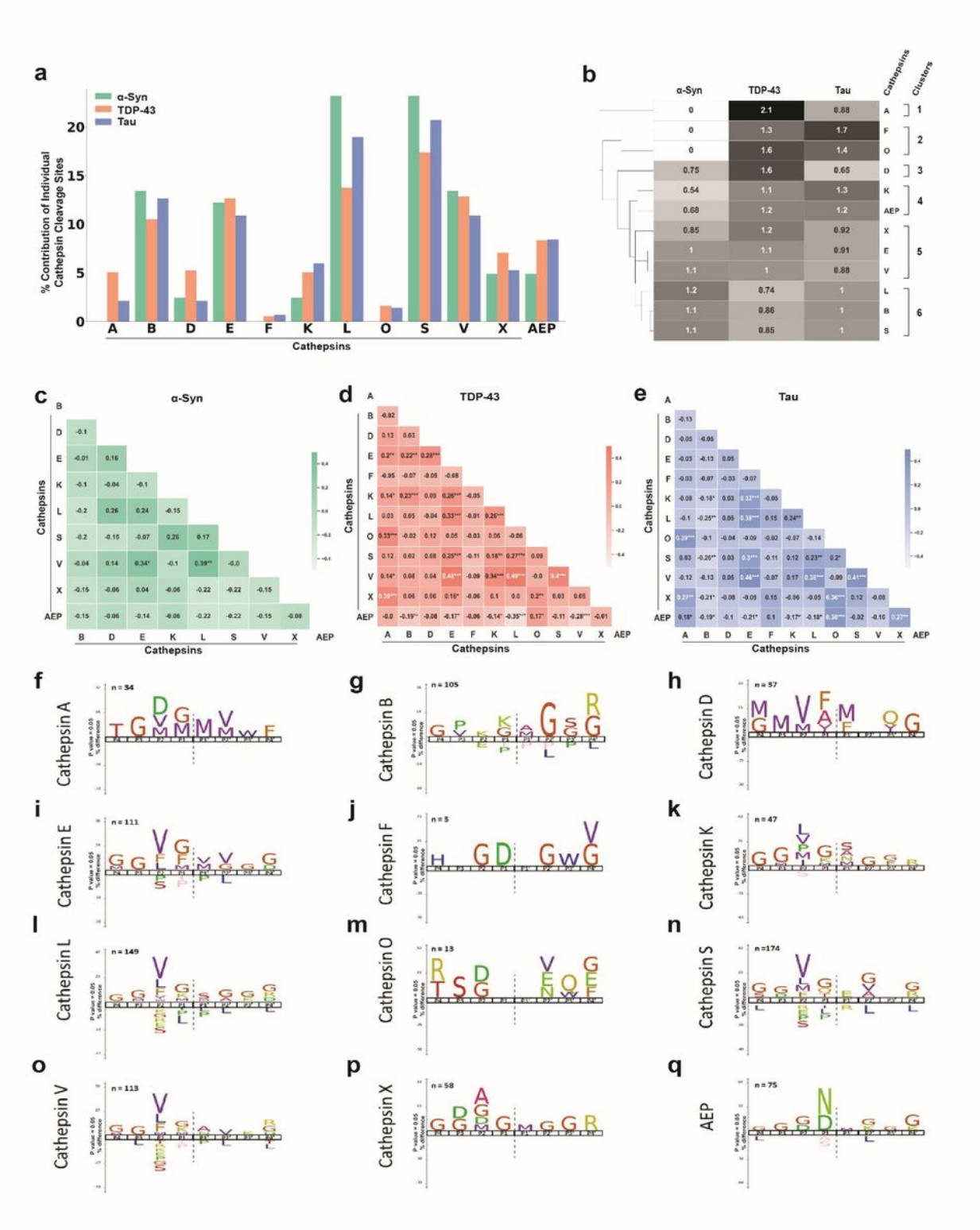

Figure 5

Lysosomal proteases exhibit distinctive abilities to process α-syn, TDP-43 and tau. a, Comparison of the number of cleavage sites within α-syn, TDP-43, and tau for each lysosomal protease relative to total number of cleavage sites. **b**, Hierarchical clustering analysis demonstrating the relative affinity of proteases for α-syn, TDP-43 or tau. Clusters identified are designated to the right. **c-e**, Pairwise correlation analyses with significance values (p-values) of unique versus redundant activity between lysosomal

proteases for α-syn (C), TDP-43 (D) and tau (E). A positive score suggests more correlation and a negative score lower correlation between protease cleavage sites. **f-q**, The iceLogo output for each of the serine (F), aspartyl (G-H) and cysteine (I-Q) proteases demonstrating the frequency of particular amino acids at the P4 – P4' positions of each protease recognition motif within α-syn, TDP-43 and tau. Amino acids that were more frequently seen are above the horizontal axis and those that were less frequently seen are below the horizontal axis. The cleavage site is indicated with a vertical hatched line. *p<0.05, **p<0.01, or ***p <0.001.

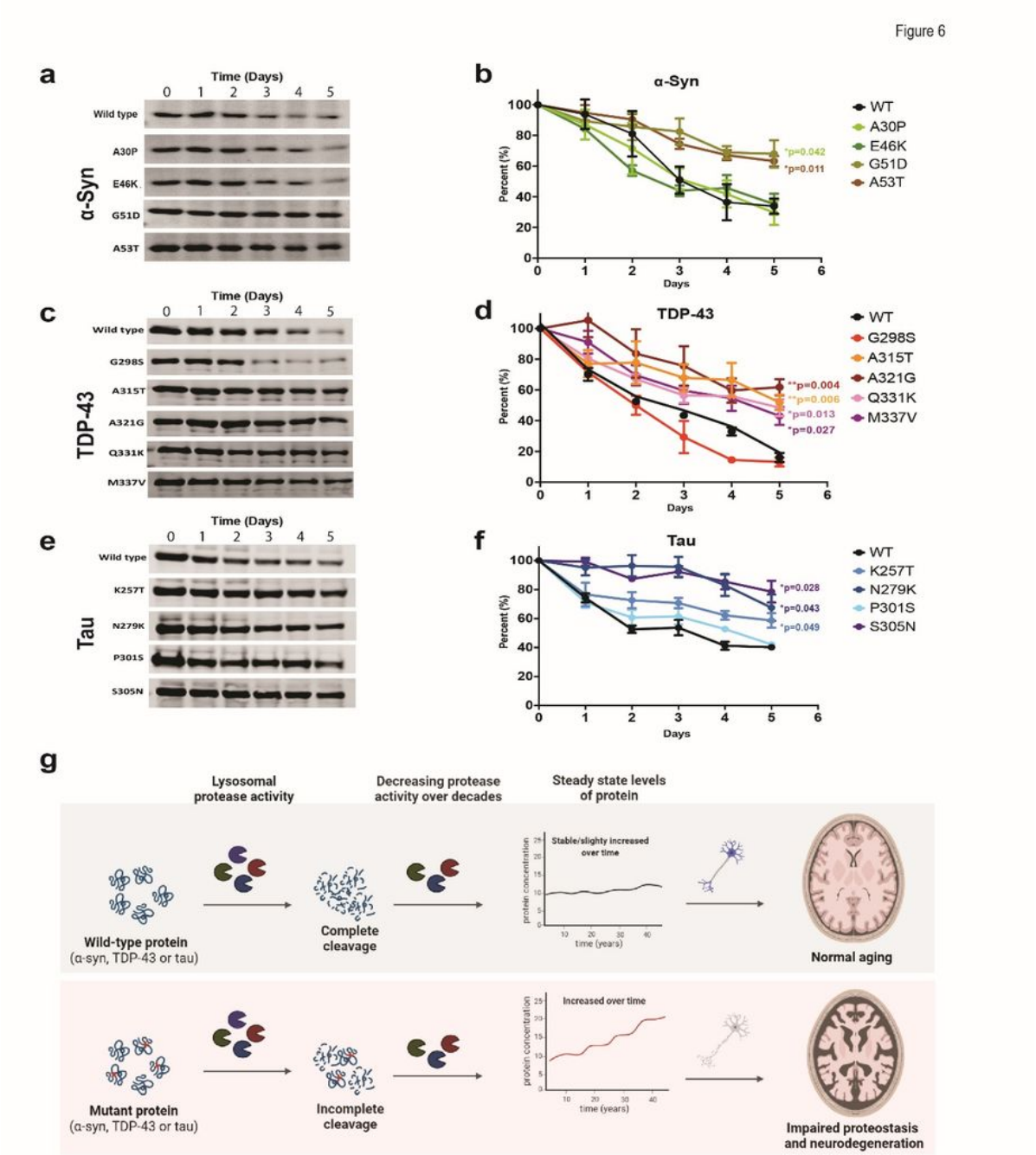

Figure 6

Pathogenic mutations prolong α-syn, TDP-43 and tau half-life. a-f, Differentiated SH-SY5Y cells expressing inducible, FLAG-tagged WT or mutant α-syn (A-B), TDP-43 (C-D) or tau (E-F) were treated for 24 hours with doxycycline. Antibiotic was removed (t=0 days), lysates were collected at each time point as indicated and samples underwent SDS-PAGE and anti-FLAG Western blotting. Representative images demonstrating clearance of WT or mutant protein over time are shown in A, C, and E with quantification of three independent replicates shown in B, D and F. **g**, Proposed model for how mutations in α-syn, TDP-43 and tau can gradually increase steady state levels of protein over decades to predispose to impaired proteostasis, protein aggregation and neurodegeneration. *p<0.05, **p<0.01, or ***p <0.001.

Supplementary Files

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

- SupplementaryData1.xlsx
- SampognaroAryaExtendedDataandSupplementaryMaterialsNature.docx