DEGRONOPEDIA: a web server for proteome-wide inspection of degrons

Natalia A. Szulc [1,*], Filip Stefaniak [2], Małgorzata Piechota [1], Anna Soszyńska [1], Gabriela Piórkowska[1], Andrea Cappannini[2], Janusz M. Bujnicki [2], Chiara Maniaci [3] and Wojciech Pokrzywa [1,*]

[1]Laboratory of Protein Metabolism, International Institute of Molecular and Cell Biology in Warsaw, 4 Ks. Trojdena Str., 02-109 Warsaw, Poland
[2]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 4 Ks. Trojdena Str., 02-109 Warsaw, Poland
[3]Medical Research Council (MRC) Protein Phosphorylation and Ubiquitylation Unit, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK

*To whom correspondence should be addressed. Tel: +48 225970743; Email: wpokrzywa@iimcb.gov.pl
Correspondence may also be addressed to Natalia A. Szulc. Email: nszulc@iimcb.gov.pl

Abstract

E3 ubiquitin ligases recognize substrates through their short linear motifs termed degrons. While degron-signaling has been a subject of extensive study, resources for its systematic screening are limited. To bridge this gap, we developed DEGRONOPEDIA, a web server that searches for degrons and maps them to nearby residues that can undergo ubiquitination and disordered regions, which may act as protein unfolding seeds. Along with an evolutionary assessment of degron conservation, the server also reports on post-translational modifications and mutations that may modulate degron availability. Acknowledging the prevalence of degrons at protein termini, DEGRONOPEDIA incorporates machine learning to assess N-/C-terminal stability, supplemented by simulations of proteolysis to identify degrons in newly formed termini. An experimental validation of a predicted C-terminal destabilizing motif, coupled with the confirmation of a post-proteolytic degron in another case, exemplifies its practical application. DEGRONOPEDIA can be freely accessed at degronopedia.com.

Graphical abstract
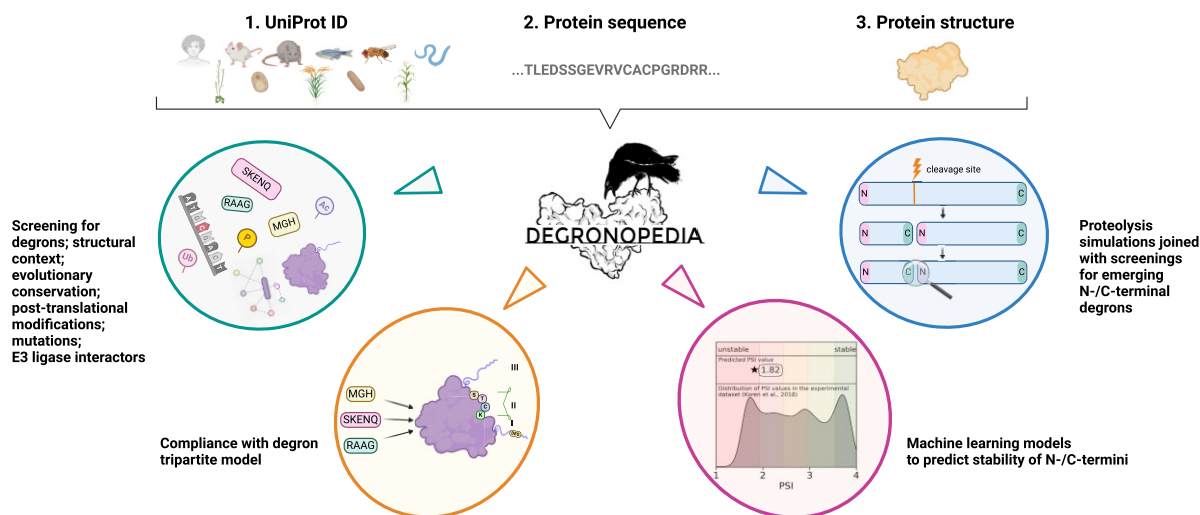
Introduction

In the ubiquitin-proteasome system (UPS), E3 ubiquitin ligases are critical for tagging proteins for degradation, primarily attaching ubiquitin to their lysine residues ([1],[2]), with emerging evidence indicating cysteine, serine, and threonine residues also acting as ubiquitination sites ([3],[4]). Proteins are targeted for degradation through degrons, typically short linear motifs found in disordered regions, which can be constitutive or conditional, the latter emerging from post-translational modifications (PTMs) ([5]). It is important to emphasize that degrons

---

do not solely exist as linear sequences. For example, numerous proteins possess zinc fingers and a structurally conserved β-hairpin loop, potentially serving as structural degrons (6,7). Degrons are not restricted in location but those at the amino or carboxyl termini, integral to the N- or C-degron pathways, have been extensively studied (5,8–14).

Recent research suggests that ubiquitination and subsequent degradation depend not just on degron recognition. The tripartite degron model proposed by Guharoy *et al.* introduces secondary degrons, comprising lysine residues near the primary degron, and tertiary degrons, which are flexible, intrinsically disordered regions (IDR) nearby, facilitating protein unfolding before proteasomal entry (15). The secondary and tertiary degrons are suggested to play subsidiary roles that affect ubiquitin-signaling. Deficiency of one of the elements of the tripartite degron model, e.g. an IDR near a ubiquitinated lysine, can result in non-proteolytic ubiquitination functions (15).

While research on degron motifs and their physiological roles is growing, there are few bioinformatics tools for degron motif screening and analysis. The APC/C degron repository (16) offers insights into APC/C degron sequence determinants, including information on disordered regions and post-translational modifications. The Eukaryotic Linear Motif (ELM) resource (17) allows detection of 32 degron motifs in proteins, providing structural context. An interactive web table based on ELM data and further manual curation, released in 2017, lists degron motifs (https://dosztanyi.web.elte.hu/CANCER/DEGRON/TP.html) (18). Additionally, deep learning models like Degpred (19) and deepDegron (20) have been developed for predicting degrons and their disruption by mutations, respectively. However, deepDegron, which predicts the potential for a protein sequence to contain a degron, is a stand-alone tool requiring specific inputs, like mutations in a Mutation Annotation Format file. Degpred, while limited to the human proteome, does not provide PTMs information that can affect degron functionality and mainly predicts internal degrons based on sequence alone, without considering their tertiary protein structure. DegronMD (21), a new resource, predicts and maps human internal degrons as well; however, its training was based on a limited dataset of 23 internal degron motifs from the ELM database. To our knowledge, no existing resource comprehensively compiles all known degron motifs for thorough screening in query proteins, covering both termini and internal areas, nor provides detailed data on degron site context, such as in the proposed tripartite degron model, or offer options for customized calculations.

Addressing these complexities, we introduce DEGRONOPEDIA, a web server for identifying and analyzing degron motifs in proteins. It provides detailed analyses of sequence, spatial, and tripartite model context, including PTMs and mutations, and features pre-calculated multiple sequence alignments (MSA) for evolutionary insights, with an option for custom uploads. Moreover, DEGRONOPEDIA allows for the simulation of sequence cleavage and identification of potential N-/C-degrons post-proteolysis. It also employs machine learning (ML) models to predict protein terminal stability, trained on datasets from high-throughput human proteome terminal stability studies.

Our successful identification of a degron at the C-terminus of FBXL15, as well as confirming the presence of a predicted degron in SDE2 post-proteolysis, demonstrates DEGRONO-

PEDIA usefulness for researchers in many areas. Our server offers a user-friendly interface, comprehensive tutorial, documentation and introductory video, making it accessible to a wide range of users. To the best of our knowledge, DEGRONOPEDIA is unique in its scope by providing the most current and extensive collection of degrons, coupled with detailed analyses of their molecular characteristics, ML-based predictions, and simulations of protein cleavage.

Materials and methods

Inputs

Three input types can be used for querying DEGRONOPEDIA: (i) a UniProt ID of a protein from the reference proteome (according to the UniProt database (22)) of one of the selected model organisms—*H. sapiens, M. musculus, R. norvegicus, D. rerio, D. melanogaster, C. elegans, S. cerevisiae, S. pombe, A. thaliana, O. sativa* or *Z. mays*, (ii) a protein sequence in the FASTA format and (iii) a protein structure (either a protein model or a structure obtained in experimental studies) in the PDB format; any submitted PDB file must contain a protein monomer with only one model and one chain with continuous numbering starting from 1 and not exceeding 5 MB in size. Regardless of the input type, the query protein must be below 40 000 canonical amino acids.

Outputs

DEGRONOPEDIA output varies depending on the type of input: a query by UniProt ID delivers the most comprehensive degron information, including a full tripartite model, N-/C-terminus stability expressed as Protein Stability Index (PSI) based on the Global Protein Stability (GPS) studies (11,12) (experimental and/or predicted values), data on evolutionary conservation, structural context (derived from corresponding AlphaFold2 model from the AlphaFold Protein Structure database (23,24)), PTMs, mutations, physiologically-relevant proteolytic cleavage sites, and E3 ubiquitin ligase interactors. In contrast, a query by FASTA provides basic screening output, missing experimental data. A query by structure offers a moderate level of detail with the tripartite model but lacks experimental data. Lastly, combining structure with UniProt ID yields the same extensive information as a UniProt ID query, except that the structural data, i.e. secondary structure, solvent-accessibility, IDRs, are based on the uploaded structure. Irrespective of the query type, users can download an xlsx file with all the results, organized into distinct sheets. The summary of DEGRONOPEDIA workflow and the granularity of the output depending on the input type is presented in Figure 1.
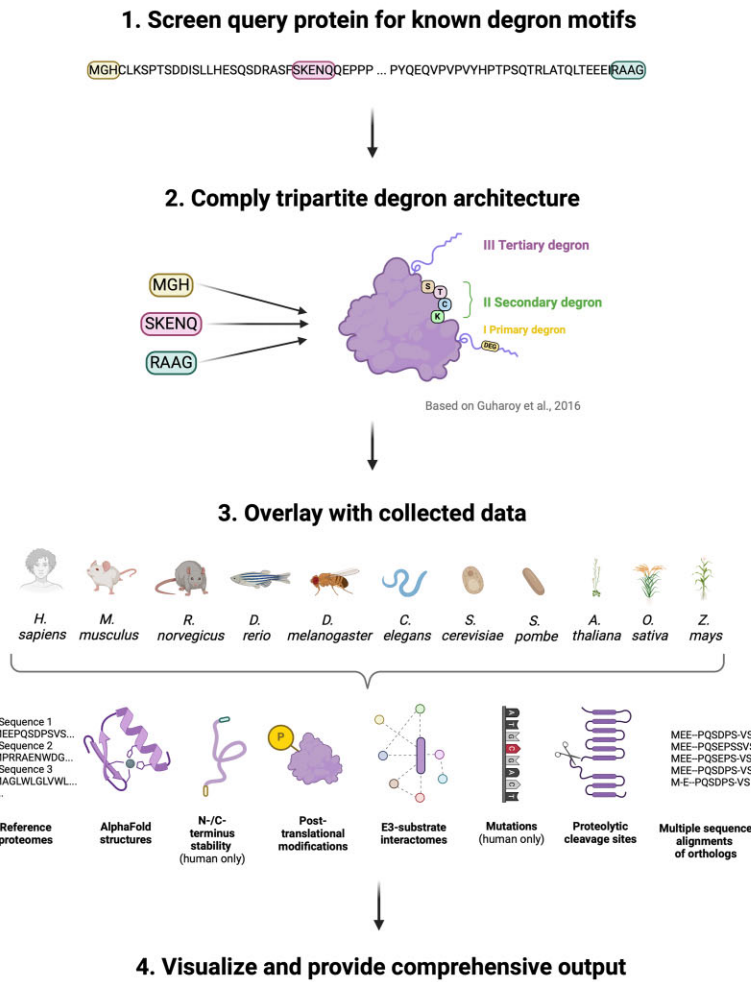
Implementation

Datasets

Degron motifs. We compiled over 1800 degron motifs (5,11,12,17,25–31). Each motif was defined as either N-terminus, C-terminus or internal, regarding its location. The list of degron motifs is continuously updated and can be downloaded from the server.

Reference proteomes and structural data. Reference proteomes for *H. sapiens, M. musculus, R. norvegicus, D. rerio, D. melanogaster, C. elegans, S. cerevisiae, S. pombe, A. thaliana, O. sativa* and *Z. mays* were obtained from the UniProt

A

1. Screen query protein for known degron motifs

MGHCLKSPTSDDISLLHESQSDRASFSKENQQEPPP ... PYQEQVPVPVYHPTPSQTRLATQLTEEEIRAAG

2. Comply tripartite degron architecture

MGH
SKENQ
RAAG

III Tertiary degron
II Secondary degron
I Primary degron

Based on Guharoy et al., 2016

3. Overlay with collected data

H. sapiens *M. musculus* *R. norvegicus* *D. rerio* *D. melanogaster* *C. elegans* *S. cerevisiae* *S. pombe* *A. thaliana* *O. sativa* *Z. mays*

>Sequence 1
MEEPQSDPSVS...
>Sequence 2
MPRRAENWDG...
>Sequence 3
MAGLWLGLVWL...
....

Reference proteomes | AlphaFold structures | N-/C-terminus stability (human only) | Post-translational modifications | E3-substrate interactomes | Mutations (human only) | Proteolytic cleavage sites | Multiple sequence alignments of orthologs

MEE~PQSDPS-VS
MEE~PQSEPSSVS
MEE~PQSEPS-VS
MEE~PQSDPS-VS
M-E-PQSDPS-VS

4. Visualize and provide comprehensive output

B

	Degron motifs occurrence ...TRLATQLTEEEIRAAG		Degron flanking regions ...RAAG...	Structural context			Tripartite degron model		Evolutionary conservation		PTMs	Mutations	Cleavage simulations			N-/C-terminus stability data (PSI)	
	user-defined	from database		Solvent-accessibility	Secondary structure	IDRs	Close in sequence	Close in structure	user-uploaded MSA	pre-calculated MSA			user-defined sites	experimental sites	predicted sites	experimental	ML-predicted
Query by UniProt ID / structure + UniProt ID	✓	✓	Defined in sequence and structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Query by sequence	✓	✓	Defined in sequence	✗	✗	✓	✓	✗	✓	✗	✗	✗	✓	✗	✓	✗	✓
Query by structure	✓	✓	Defined in sequence and structure	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✗	✓	✗	✓

Figure 1. DEGRONOPEDIA implementation details. (**A**) Operational scheme of server functioning when querying by UniProt ID. (**B**) Comparison of the output information obtained upon different query types.

database (IDs: UP000005640, UP000000589, UP000002494, UP000000437, UP000000803, UP000001940, UP000002311, UP000002485, UP000006548, P000059680, UP000007305, respectively), along with structural models from the AlphaFold Protein Structure database. In cases of proteins exceeding 2700 amino acids, an in-house script using Biopython (32) merged AlphaFold2 model fragments into single PDB files.

Orthologs data. Pre-calculated MSAs of predicted orthologs were obtained from the eggNOG5 database (33) at various taxonomic levels (Supplementary Table S1).

N-/C-termini stability data. The N-/C-termini stability data were obtained from the GPS assays (11,12). These data cover the stability of N-/C-terminal 23-mers of 48251 (variants with N-terminal methionine and without it) and 22564 human proteins, respectively, and are represented as PSI,

where its lowest values indicate the most unstable, thus possibly containing a degron motif, peptides.

Post-translational modification data. The PTMs datasets were obtained from the iPTMNet (34), PhosphoSitePlus (35), Plant PTM Viewer (36), and the PLMD (37) databases as well as from the literature, where we manually compiled datasets of non-canonical ubiquitination and arginylation (references are provided within each reported PTM).

Mutation data. Human mutation data, specifically 'Substitution - Missense' types, were obtained from the COSMIC database (38).

E3 interactome data. The interactomes were acquired from the BioGRID (39), IntAct (40), UbiNet 2.0 (41) databases and from the literature (11,26), focusing on E3 ligases interactions. E3 ligase annotations were sourced from the AMIGO web server (using query GO:00616300) (42), ESBL (43) and Ubi-Net 2.0.

Proteolytic cleavage sites data. The experimental proteolytic cleavage sites specific for a given protein with adherent information about the involved proteolytic enzymes were derived from the MEROPS database (44) and literature (references are provided within each reported cleavage site). The MEROPS dataset was subsequently filtered to contain only cleavage sites classified as physiologically relevant with present information about their exact position in the sequence.

Secondary structure and solvent accessibility

The secondary structure and solvent-accessible surface area of each residue are calculated by the server using the mkdssp software (45,46), with the latter being normalized to the relative solvent-accessibility (RSA) by the Sander method (47).

Evolutionary conservation score

For every available MSA, four different degron conservation scores are computed, reflecting the motif preservation throughout the MSA and its adjacent area (that can be adjusted), as illustrated in Supplementary Figure S1.

Disorder prediction

DEGRONOPEDIA offers two methods to predict disordered regions in proteins: (i) using pLDDT/LDDT (predicted Local Distance Difference Test/Local Distance Difference Test, respectively) values from AlphaFold2/RoseTTAFold (48) models for structure-containing queries and (ii) using IUPred3 software (49), applicable to all query types, which calculates a disorder score for each amino acid based on the query sequence. If choosing the former, a structure model from AlphaFold2/RoseTTAFold is required with a B-factor column in the PDB file filled with valid pLDDT or LDDT scores, respectively. Both methods have adjustable thresholds (default <70 for the former and >0.5 for the latter to consider a residue as disordered) for defining disordered residues on the server, and IDRs are identified as continuous regions meeting the set length.

Proteolysis prediction

Cleavage analysis of a query protein can be conducted based on a user-specified cleavage motif/site, MEROPS dataset, or through predictions made using the Pyteomics module (50). This module forecasts cleavage sites for 35 distinct proteolytic enzymes, implementing the cleavage prediction rules of the PeptideCutter Expasy web server (51).

Machine learning models development

To develop our ML models, we used experimental N-/C-termini stability data, expressed as PSI values, from 23-mers covering the human proteome, considering both presence and absence of the initial methionine residue in N-termini. The datasets were divided into training and testing sets (90:10 ratio), and CatBoost regression models (52) were trained for each terminus (C-termini, N-termini with and without the initial methionine residue), optimizing hyperparameters with Optuna framework (53) and five-fold cross-validation (using random permutations cross-validation implemented in the scikit-learn Python library (54), with a 20% validation set). Descriptors included peptide sequences, RDKit descriptors (RDKit: Open-source cheminformatics. https://www.rdkit.org), Gravy hydrophobicity index, and the Peptides module (https://github.com/althonos/peptides.py). All descriptors were calculated for the whole sequence, and the first (excluding the N-terminal methionine regardless of whether we considered it as present or absent) or the last (for the C-terminus) ten, eight, six, four and two amino acids. Model performance was evaluated using a coefficient of determination (R^2) and root mean square error (RMSE). The performance of all three models was similar, reaching $R^2 = 0.796$ and RMSE $= 0.495$ for the N-terminus with initiator methionine cleaved, $R^2 = 0.812$ and RMSE $= 0.443$ for the N-terminus with initiator methionine not cleaved, and $R^2 = 0.815$ and RMSE $= 0.316$ for C-terminus (Figure 2). The server visualizes predicted PSI values, mapping them to experimental data distributions and classifying them into stability categories using quantile thresholds. Neptune management system was used for tracking ML experiments (neptune.ai: experiment tracker; https://neptune.ai).

Customizable parameters

Screening for known degron motifs is dynamically performed on the web server, considering ten parameters the user can customize on the homepage. In short, they describe thresholds related to the distinct regions of the tripartite degron architecture, classifying residues as buried or disordered and calculating evolutionary conservation (Supplementary Table S2).

Visualizations

Degrons visualizations in the sequence are created using FeatureViewer (55), while ProSeqViewer (56) and Logomaker (57) package are employed to display degrons within the MSAs. Graphics were created in BioRender.com.

FBXL15 case study experimental approach

Cell culture

Flp-In 293 HEK293 cell line (female embryonic kidney epithelial cell line; ThermoFisher Scientific; authentication certificate issued by the manufacturer—checked for viability, mycoplasma, sterility and b-galactosidase activity) were cultivated in Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 10% heat-inactivated Fetal Bovine Serum (Sigma) and 1% Antibiotic-Antimycotic (Gibco) at 37°C, 5% CO_2 in a humidified incubator.

Plasmids

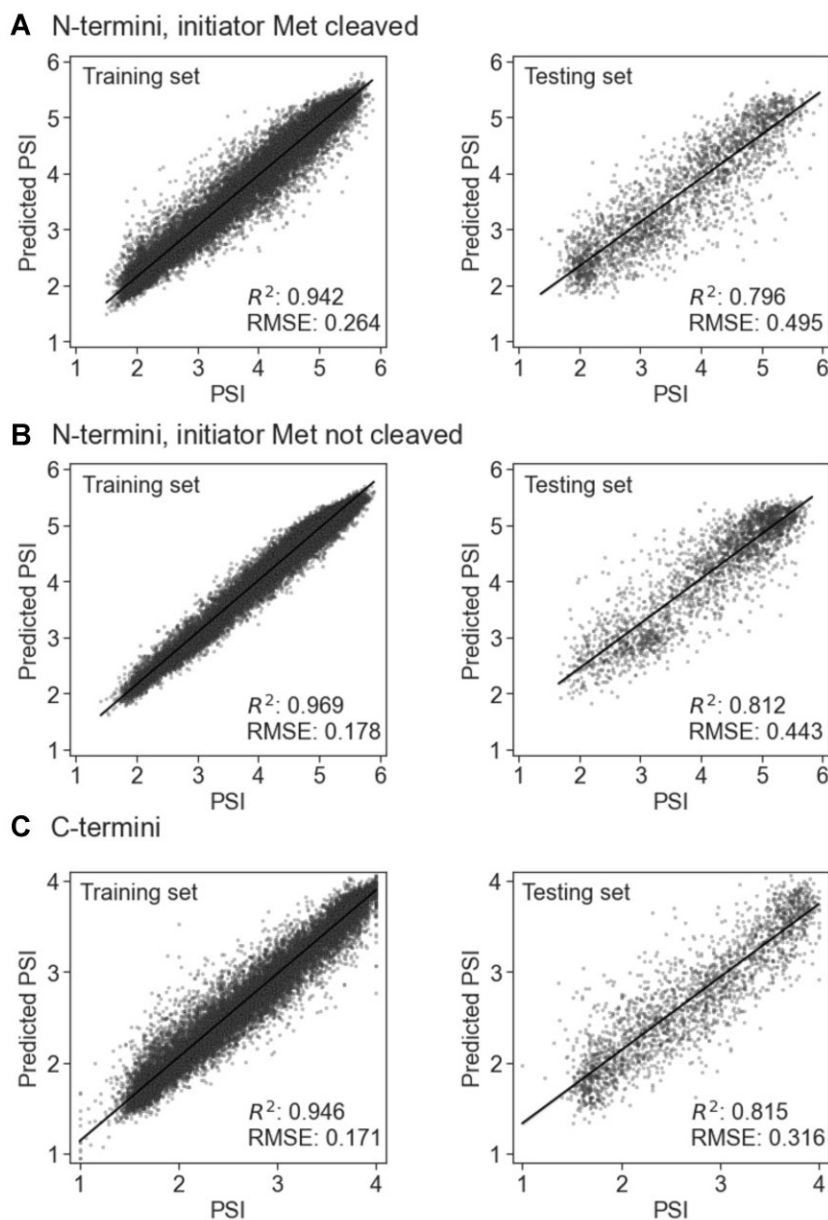HiBiT fusion vectors were constructed by the sequence and ligation independent cloning (SLIC) method (58)

Figure 2. Predictions of the PSI using ML CatBoost regression models for N-termini with the initiator methionine (Met) cleaved (**A**) or not cleaved (**B**) and for C-termini (**C**). Scatter plots show a regression line with 95% confidence.

using the parental vector pBiT3.1-N (Promega) or pBiT3.1-C (Promega) for the N- or C-terminal fusion variant, respectively, linearized with EcoRI (N-terminal variant) and SacI (C-terminal variant) enzymes, dephosphorylated and mixed with human FBXL15 sequence PCR-amplified from HEK293 cDNA (FBXL15 protein sequence corresponding to UniProt ID: Q9H469).

Cycloheximide chase assay

Cells preparation and transfection. HEK293 cells were seeded in white 96-well tissue culture plates (Greiner) at a density of 10.000 cells in a total volume of 100 µl per well. After 38–40 h, cells were transiently transfected with 2.5 ng HiBiT-tagged human FBXL15 expression constructs, diluted in carrier DNA (Promega) to obtain a final DNA amount of 50 ng/well. Transfection was carried out using the FuGENE HD

Transfection Reagent (Promega) according to the manufacturer's guidelines in the Opti-MEM I Reduced-Serum Medium (Gibco), maintaining the 3:1 FuGENE HD:DNA ratio. Cells were incubated for 23 h at 37°C, 5% CO_2.

Cycloheximide assay. Cycloheximide (CHX; BioShop) was added to wells at a final 50 µg/ml concentration for 4 h. Treatment was performed in triplicate. Where indicated, cells were concomitantly treated with 20 µM MG132 (Selleck Chemicals) for 4 h. After a 4-h incubation with CHX, the Nano-Glo HiBiT Lytic Detection Assay (Promega) was performed according to the manufacturer's guidelines. Briefly, 120 µl of the Nano-Glo HiBiT Lytic Reagent was added to each well. Samples were mixed on an orbital shaker for 5 min and incubated for 10 min before luminescence measurement using TECAN Infinite 200 Pro plate reader equipped with the Magellan Pro software and an integration time of 1000 ms. The

luminescence measurements were normalized in each well to the number of living cells (see below).

Cell viability assay. To assess cell viability during the CHX assays, they were multiplexed with the CellTiter-Fluor Cell Viability Assay (Promega) following the manufacturer's guidelines. Briefly, after 3-h incubation with CHX, 20 µl of the 5× concentrated CellTiter-Fluor Reagent was added to all wells. Cells were incubated at 37°C for 1 h, and fluorescence was measured using the Tecan Infinity M1000 fluorescence plate reader equipped with the Magellan Pro software with the parameters setup of 390nmEx/505nmEm. The untransfected cells were used as the global viability reference.

SDE2 case study experimental approach
Cell culture
HEK293 cell line (ATCC; authentication certificate issued by the manufacturer—checked for viability, mycoplasma, sterility and b-galactosidase activity) were cultivated in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% heat-inactivated fetal bovine serum (Sigma) and 1% antibiotic–antimycotic (Gibco) at 37°C, 5% CO_2 in a humidified incubator.

Plasmids
For transient transfection, the following were cloned into pCMV3 plasmid: HA-SDE2-FLAG (DU75320), HA-SDE2-FLAG G76,77A (DU75421), HA-SDE2-FLAG G76A (DU75344), HA-SDE2-FLAG G77A (DU75364), HA-SDE2-FLAG K78M (DU75354), HA-SDE2-FLAG K78A (DU75377). All constructs were sequence-verified by the DNA Sequencing Service, University of Dundee (https://www.dnaseq.co.uk). These constructs are available to request from the MRC PPU Reagents and Services webpage (http://mrcppureagents.dundee.ac.uk) and the unique identifier (DU) numbers provide direct links to the cloning strategies and sequence details.

Cycloheximide chase assay
Cells preparation and transfection. HEK293 cells were transferred in six-well plates with 6×10^5 cells per well in 2 ml media in order to achieve 80% confluence the day after. Cells were transfected with 0.020 µg/µl plasmid solution in the Opti-MEM I Reduced-Serum Medium (Gibco) using FuGENE HD Transfection Reagent (Promega). All cells were lysed 48 h after transfection.

Cycloheximide assay. For the CHX chase experiment, 0.1% v/v of DMSO CHX (BioShop; 100 mg/ml final concentration) was added to the desired well at different time points—six, four, two, one and half hours before lysis. For the proteasome inhibitor treatment, MG132 (S2619, Selleck Chemicals) was added into the desired wells at 50 µM final concentration and 0.1% v/v of DMSO. DMSO (0.1% v/v final concentration) was added to the remaining wells as vehicle control. Cells were incubated for 4 h before harvesting. After incubation, the medium was removed, and cells were washed twice with ice-cold phosphate buffer saline. RIPA lysis buffer (Sigma) containing cOmplete protease inhibitor cocktail (Roche) was added, and the cells were collected by scraping. After the removal of the insoluble fraction by centrifugation, the protein concentration of the clarified lysate was determined by the Pierce Coomassie (Bradford) Protein Assay Kit.

Immunoblotting. Protein extracts were fractionated by SDS-PAGE on 4–12% Tris-acetate NuPage Novex (Life Technologies) polyacrylamide gels and transferred to a nitrocellulose membrane using wet transfer. The membrane was then blocked with 5% w/v bovine serum albumin in Tris-buffered saline with 0.1% w/v Tween-20. For detecting proteins, the following primary antibodies in the given concentrations were used: anti-GAPDH (Cell Signaling Technology, 14C10) 1:2000, anti-FLAG (Cell Signaling Technology, D6W5B) 1:1000, anti-HA (Cell Signaling Technology, C29F4) 1:1000. Following incubation with Licor 926-32211 IRDye 800CW Goat anti-Rabbit IgG (H + L), the signal was developed using Licor Odyssey CLx Imager.

Results
Workflow
We outline the DEGRONOPEDIA workflow for queries using UniProt IDs, which offer the most exhaustive results (see Figure 1). Other query types follow a similar process but lack certain analyses due to limited access to experimental data. The workflow begins by screening the queried protein for degron motifs from our curated dataset, followed by calculating the Gravy hydrophobicity index (59) of the protein's terminal 15 amino acids, a step informed by the significant role hydrophobicity plays in N-/C-degron recognition by various E3 ubiquitin ligases (60,61). Subsequently, DEGRONOPEDIA offers experimental N-/C-terminus stability data, represented as PSI values (specifically for human proteins), sourced from GPS studies (11,12). It also facilitates predictions using our pretrained ML models. Since these models were developed using human PSI values, we suggest their use primarily for queries related to higher mammals. The server then visualizes PSI values, categorizing them into stability classes; we also offer a standalone tool for high-throughput PSI prediction (available at https://github.com/filipsPL/degronopedia-ml-psi). Additionally, we have integrated pre-calculated MSAs and offer the option for users to submit custom MSAs, enabling the provision of evolutionary conservation scores derived from these alignments. Solvent-accessibility, location within an IDR region, or lack of secondary structure are premises of a site acting as an actual degron (15). Our server predicts IDRs and degrons' disorder using AlphaFold2/RoseTTAFold models or IUPred3 software, depending on the input type, and calculates secondary structure and relative solvent-accessibility for each degron motif. As degron sites may undergo various PTMs, with the primary role of phosphorylation, which can modulate their exposure (62), the server reports experimentally validated PTMs (up to 32 types) occurring within each found degron motif and its flanking regions. Amino acid substitutions in degron motifs can lead to altered protein stability, contributing to severe diseases such as cancer (18) or neurodegeneration (63), indicating critical sites for proper protein function. Therefore, our tool provides information about known missense mutations within the degron motifs and within PTMs in their flanking regions. For each identified degron, the server also evaluates its context in the tripartite degron model. In particular, the server provides information on solvent accessibility, secondary structure, location within an IDR, mean disorder, PTMs, and mutations for each potentially ubiquitinated residue (secondary degron; in our implementation, these are not only lysines (K) but also
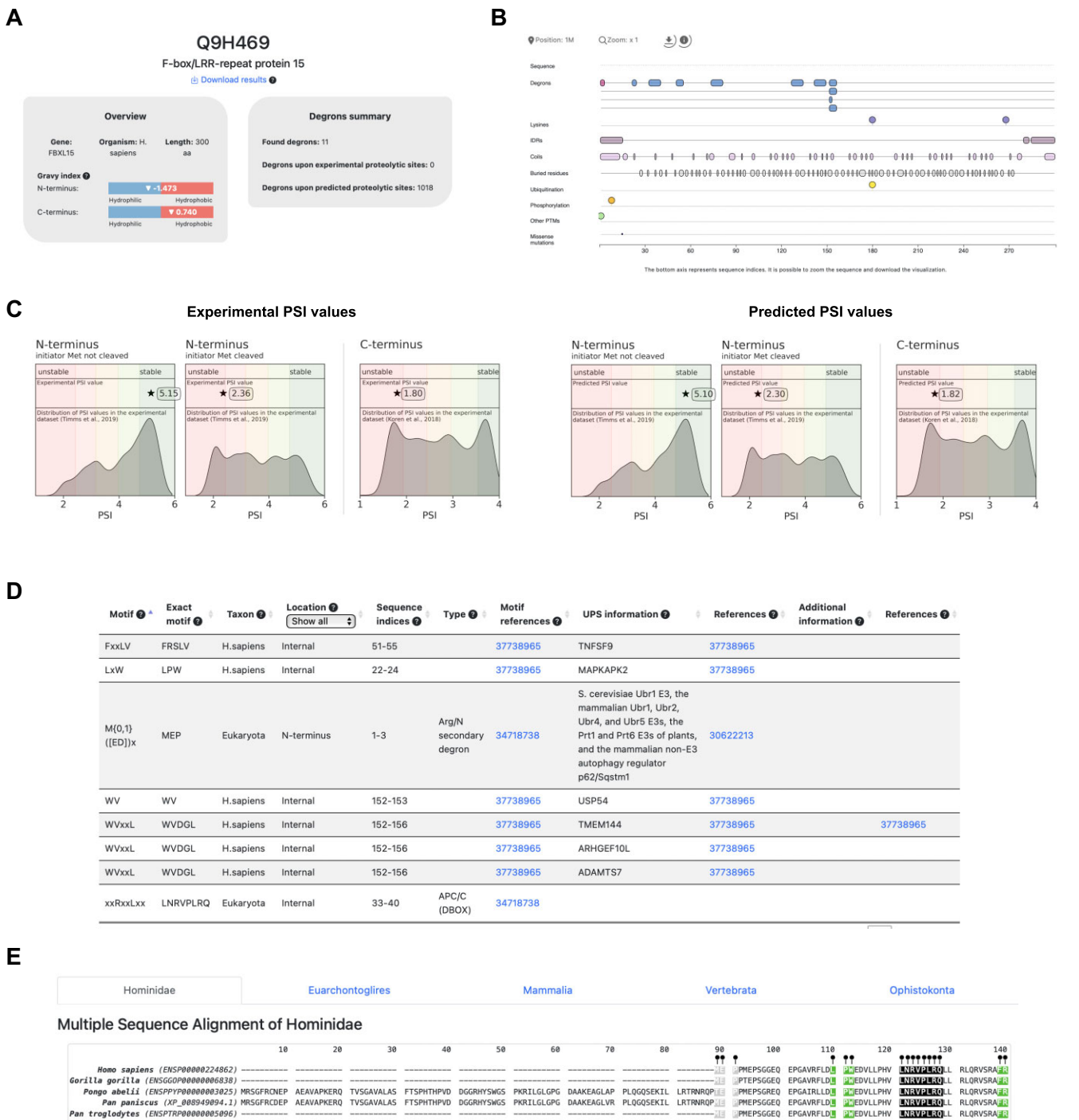
Figure 3. Overview of the DEGRONOPEDIA outputs using FBXL15 cullin-RING substrate receptor. (**A**) Overview panel with general information on the FBXL15 protein (left) and degrons summary panel with information on the number of found degron motifs (right). (**B**) Visualization of the found degron motifs, structural data, PTMs and mutations on the protein sequence. (**C**) Experimental (left) and ML-predicted (right) values of the N-/C-terminus of FBXL15 plotted on the distribution of experimental stability datasets. (**D**) Example of interactive output table providing information on the degrons found in the FBXL15. (**E**) Visualization of MSA with denoted positions of found degrons in the orthologs.

cysteines (C), serines (S) or threonines (T)) located within the degron flanking regions. Finally, the server reports the closest IDR (tertiary degron) to each of the secondary degrons. It has been shown that protein turnover may be regulated by proteolytic enzymes that cleave the protein, leading to new N- and C-termini, which may act as degrons (5). Thus, DEGRONO-PEDIA simulates protein cleavage based on user-defined mo-

tifs, validated cleavage sites, or predicted sites for various proteases, subsequently screening the cleavage products for degrons. The final information provided relates to E3 ligases known to interact with the queried protein. The occurrence of degrons in the sequence, along with their unique features, is visualized, and the outcomes are presented in easy-to-navigate interactive tables. The tutorial and introductory video, acces-
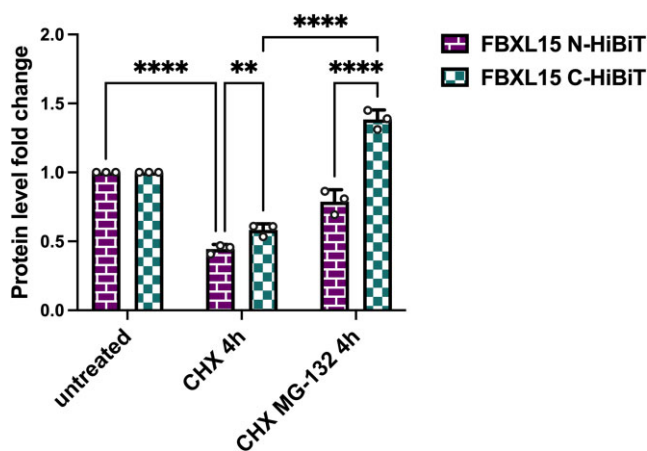
Figure 4. FBXL15 possess a destabilizing C-terminus. N- or C-terminus-HiBiT-tagged FBXL15 turnover was measured by CHX assay with transient expression of FBXL15 fusion variants and co-treatment with 50 μg/ml CHX for 4 h and, where indicated, 20 μM MG132 proteasome inhibitor. Protein levels were measured and normalized to the number of living cells as described in Materials and Methods; assays were normalized to the corresponding measurement for each protein variant from control time 0′. Error bars denote the standard deviation from the mean derived from three biological replicates; dots represent the biological replicates. Each biological replicate is a mean of three technical replicates. Data was analyzed using two-way ANOVA and the significance levels obtained from Tukey's multiple comparisons test are indicated for the compared conditions (**** $P \leq 0.0001$; ** $P \leq 0.01$).

sible from the web server, provide clear step-by-step instructions on the server functionality. The excerpt of outputs that DEGRONOPEDIA provides is illustrated in Figure 3.

Computation times

Queries are processed relatively fast when degron conservation scores are not considered – depending on the submitted query type, the mean calculation time for a 393 aa p53 protein ranges from 25 to 27 s, for a 2549 aa mTOR from 42 to 45 s, and for a 5890 aa AHNAK from 78 to 90 s. The calculation of conservation scores requires longer times—88 s for p53, 180 s for mTOR, and 145 s for AHNAK. A comparison of computation times and different types of queries with varying input protein lengths can be found in Supplementary Table S3.

Example application: validating ML-predicted degron in FBXL15

To ascertain the accuracy of DEGRONOPEDIA in predicting protein terminal stability, our study centered on human FBXL15 (UniProt ID: Q9H469), a cullin-RING ligase complex receptor (64), identified by the tool as having an unstable C-terminus (Figure 3). To validate these predictions, which are based on ML model that has been trained on stability data of C-terminal 23-mers, we employed Promega HiBiT system, a technique previously validated by our research (65) for its efficacy in examining the stability of other cullin receptors. We attached the HiBiT tag, comprising a short 11 amino acid sequence, to both the N- and C-termini of FBXL15, aiming to obstruct the protein's C-terminal degron and potentially modify its stability, in line with DEGRONOPEDIA projections. To assess changes in stability, we conducted cycloheximide (CHX) chase assays in HEK293 cells, a methodology integral to inhibiting protein synthesis and thereby allowing the track-

ing of FBXL15 degradation. The interaction between HiBiT and the complementary LgBiT peptide facilitated the precise quantification of FBXL15 levels through luminescence. Our findings revealed a notable increase in FBXL15 stability when the HiBiT tag was attached to its C-terminus, particularly after a 4-hour CHX treatment (Figure 4). This heightened stability contrasted with the lower stability observed in the case of the N-terminal HiBiT variant. Further, the application of the proteasome inhibitor MG132 led to significant accumulation of both protein variants, especially the C-terminally tagged version, indicating proteasome-dependent degradation of FBXL15. Of note, we chose FBXL15, as the exact experimental data of 23-mer C-terminal stability data were available for this protein, which confirms the reliability of our ML predictor and affirms the presumed instability of FBXL15 untagged C-terminus, pinpointing DEGRONOPEDIA accuracy in the studies on full-length proteins.

Example application: verifying predicted degron in SDE2 post-proteolysis

To test DEGRONOPEDIA ability to identify and evaluate the stability of newly revealed N-termini of proteins that undergo proteolytic processing, we performed a case study on the splicing regulator SDE2 protein (UniProt ID: Q6IQ49). SDE2 plays critical roles in several cellular processes, including ribosomal biogenesis, telomere silencing, DNA replication, and pre-mRNA processing (66–68). SDE2 is expressed as a ubiquitin-fold precursor protein carrying an N-terminal ubiquitin-like domain. Proteolytic cleavage immediately downstream of the ubiquitin-like domain (SDE2-UBL) at the diglycine G76-G77 motif releases an activated fragment of the C-domain of SDE2 (SDE2-C). It was previously reported that destabilizing lysine 78 (K78) at the neo-N-terminus of SDE2-C directs it to proteasomal degradation (67). To verify if DEGRONOPEDIA would cope with simulating proteolytic cleavage by deubiquitinating enzymes (DUBs) and assessing the stability and presence of neo-N/C-terminus degrons, we analyzed human SDE2 in our tool. DEGRONOPEDIA indicated the previously reported processing site pointing at low stability of both SDE2 fragments (SDE2-UBL and SDE2-C) and an Arg/N primary degron in SDE2-C (Figure 5A). To confirm this, we expressed SDE2 tagged with HA (at N-terminus) and FLAG (at C-terminus) to monitor both fragments simultaneously using immunoblotting and anti-HA and anti-FLAG antibodies, respectively. The SDE2-C fragment was short-lived in the CHX chase assay, with a half-life of approximately 3 hours, in agreement with previous data (67). A double mutation of the G76-G77 cleavage site to alanine (G76,77A) completely abolished processing, resulting in a stable, full-length, non-cleavable SDE2. Importantly, mutations of K78 (K78M or K78A), which abolished the N-terminal degron, stabilized SDE2-C. To test whether the observed short half-life depended on active proteasomal degradation, cells overexpressing HA-SDE2-FLAG or G76,77A double mutant were treated with the proteasome inhibitor MG132 for 4 hours before harvesting. Proteasome inhibition led to increased levels of SDE2-UBL and SDE2-C fragments, consistent with the proteasome-dependent degradation expected for targets of the N-degron pathway (Figure 5B).

In the case studies featured, DEGRONOPEDIA demonstrates a unique capability absent in current methodologies: the prediction of N-/C-terminal degrons based on the most
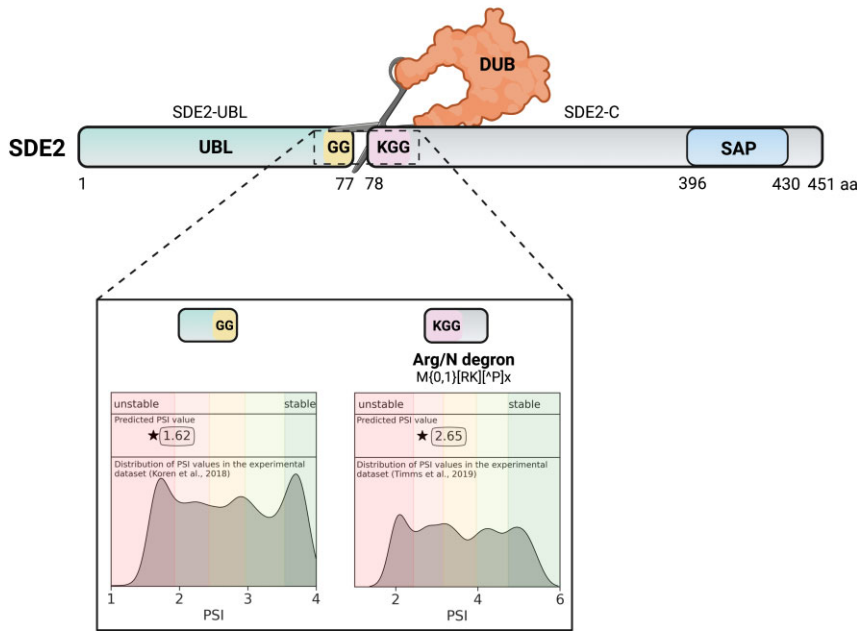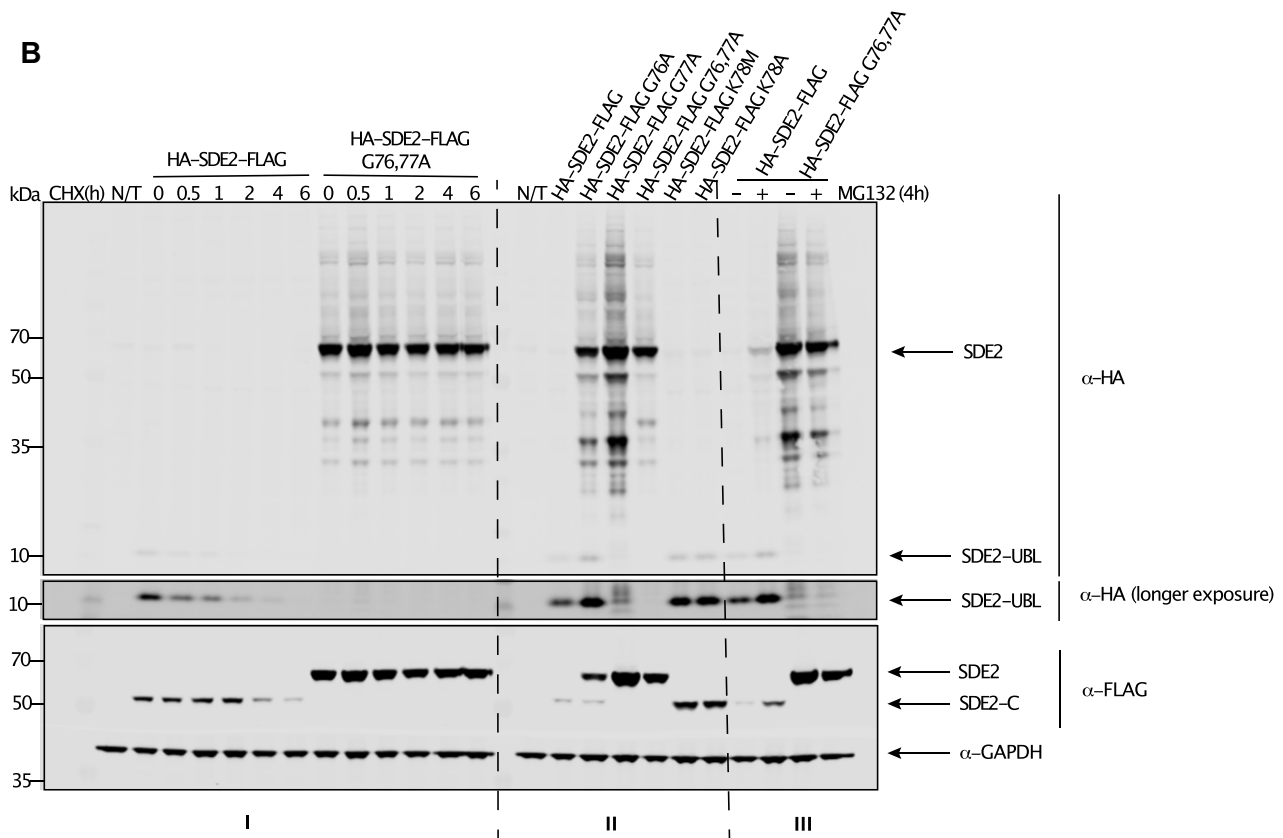
Figure 5. SDE2 proteolysis leading to new cleavage product degraded by Arg/N-degron pathway. (**A**) Schematic representation of SDE2 proteolysis occurring at the 77 amino acid position, leading to newly cleaved products. DEGRONOPEDIA-performed PSI predictions of the new termini. UBL – ubiquitin-like domain; DUB – deubiquitinating enzyme; SAP – DNA-binding SAP domain. (**B**) SDE2 stability depends on positive post-translational processing and degradation via the proteasome. I. CHX assay with transient expression of HA-SDE2-FLAG or G76,77A mutant treated. The levels of SDE2-C, SDE2-UBL and SDE2 (full-length) were determined by western blotting. II. HEK293 cells expressing indicated HA-SDE2-FLAG variants. The levels of SDE2-C, SDE2-UBL and SDE2 (full-length) were determined by western blotting. III. HEK293 cells expressing HA-SDE2-FLAG or G76,77A mutant treated with DMSO control (–) or proteasome inhibitor MG132 (+) for 4 h. The levels of SDE2-C, SDE2-UBL and SDE2 (full-length) were determined by western blotting. N/T – not treated.

comprehensive GPS data, as well as the identification of degrons emerging from proteolysis. This platform uniquely positions itself among existing online tools, a comparison of which is detailed in Supplementary Table S4.

Discussion

DEGRONOPEDIA aligns with Guharoy *et al.* tripartite degron model (15), offering an intuitive interface for comprehensive degron analysis. It accepts various input forms, including UniProt ID, FASTA sequences and PDB structures, with the depth of output tailored to the input type. The server excels in identifying known degron motifs, analyzing overlapping PTMs and mutations, and detecting novel N-/C-pathway degron motifs post-simulated proteolysis. It also evaluates evolutionary conservation scores and predicts protein terminal stability using ML models.

A key feature of DEGRONOPEDIA is its incorporation of PSI information, providing insights into terminal stability. While high PSI values do not preclude degron presence, low PSI values strongly suggest susceptibility to the N-/C-degron pathway. This is exemplified in our study on FBXL15 from the cullin-RING ligase complex, where DEGRONOPEDIA predictions of an unstable C-terminus were experimentally validated using the HiBiT system.

However, DEGRONOPEDIA has certain limitations. It does not support batch queries and can process sequences up to 40 000 amino acids or up to 5 MB file size for structure uploads. UniProt ID queries are currently restricted to proteins from 11 model organisms. For other taxa or non-reference isoforms, sequence or structure queries are the alternatives. Regarding disorder predictions, it is important to note that pLDDT/LDDT scores from AlphaFold2 models may not always align with disordered regions. Aderinwale *et al.* analysis revealed that only 30–50% of residues with low pLDDT scores matched disordered regions predicted by other methods (69). Hence, pLDDT scores should not be the sole metric for identifying IDRs. To address this, DEGRONOPEDIA allows users to adjust the pLDDT threshold for disorder or employ the IUPred3 tool for sequence-based predictions.

DEGRONOPEDIA also tackles the challenge of false positives in degron identification due to the highly degenerate nature of short linear motifs. The server enhances prediction accuracy by integrating additional contextual information, including secondary structure, RSA, IDRs, PTMs, mutations, E3 ligase interactions, and evolutionary conservation. This comprehensive approach aids in distinguishing functional degrons, offering a reliable starting point for experimental validation.

Future DEGRONOPEDIA updates will focus on improving IDR predictions, offering features for consecutive N-/C-terminus depletion analysis, and enhancing visualization through structure viewers. We plan to continually integrate new datasets to keep the server updated with the latest research developments. We also recognize the evolving understanding of degron architecture, transitioning from simple, linear motifs to multifaceted, non-linear arrangements that present significant challenges for existing predictive tools such as DEGRONOPEDIA. A substantial shift in our methodology is required to address this complexity, underscoring the critical demand for sophisticated computational techniques and comprehensive structural data to identify non-continuous degrons reliably.

Data availability

DEGRONOPEDIA is a comprehensive online web server available at https://degronopedia.com.

Supplementary data

Supplementary Data are available at NAR Online.

Conflict of interest statement

Janusz M. Bujnicki is an Executive Editor of *Nucleic Acids Research*.

References

1. Komander,D. and Rape,M. (2012) The ubiquitin code. *Annu. Rev. Biochem.*, **81**, 203–229.
2. Dikic,I. and Schulman,B.A. (2023) An expanded lexicon for the ubiquitin code. *Nat. Rev. Mol. Cell Biol.*, **24**, 273–287.
3. Squair,D.R. and Virdee,S. (2022) A new dawn beyond lysine ubiquitination. *Nat. Chem. Biol.*, **18**, 802–811.

4. Kravtsova-Ivantsiv,Y. and Ciechanover,A. (2012) Non-canonical ubiquitin-based signals for proteasomal degradation. *J. Cell Sci.*, **125**, 539–548.

5. Varshavsky,A. (2019) N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 358–366.

6. Sievers,Q.L., Petzold,G., Bunker,R.D., Renneville,A., Slabicki,M., Liddicoat,B.J., Abdulrahman,W., Mikkelsen,T., Ebert,B.L. and Thoma,N.H. (2018) Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science*, **362**, eaat0572.

7. Cowan,A.D. and Ciulli,A. (2022) Driving E3 ligase substrate specificity for targeted protein degradation: lessons from Nature and the laboratory. *Annu. Rev. Biochem.*, **91**, 295–319.

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

9. Hwang,C.-S., Shemorry,A. and Varshavsky,A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science*, **327**, 973–977.

10. Chen,S.J., Wu,X., Wadas,B., Oh,J.H. and Varshavsky,A. (2017) An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science*, **355**, eaal3655.

11. Koren,I., Timms,R.T., Kula,T., Xu,Q., Li,M.Z. and Elledge,S.J. (2018) The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell*, **173**, 1622–1635.

12. Timms,R.T., Zhang,Z., Rhee,D.Y., Harper,J.W., Koren,I. and Elledge,S.J. (2019) A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science*, **365**, eaaw4912.

13. Timms,R.T. and Koren,I. (2020) Tying up loose ends: the N-degron and C-degron pathways of protein degradation. *Biochem. Soc. Trans.*, **48**, 1557–1567.

14. Gonda,D.K., Bachmair,A., Wunning,I., Tobias,J.W., Lane,W.S. and Varshavsky,A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.*, **264**, 16700–16712.

15. Guharoy,M., Bhowmick,P., Sallam,M. and Tompa,P. (2016) Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat. Commun.*, **7**, 10239.

16. He,J., Chao,W.C., Zhang,Z., Yang,J., Cronin,N. and Barford,D. (2013) Insights into degron recognition by APC/C coactivators from the structure of an Acm1-Cdh1 complex. *Mol. Cell*, **50**, 649–660.

17. Kumar,M., Michael,S., Alvarado-Valverde,J., Zeke,A., Lazar,T., Glavina,J., Nagy-Kanta,E., Donagh,J.M., Kalman,Z.E., Pascarelli,S., *et al.* (2024) ELM-the eukaryotic Linear Motif resource-2024 update. *Nucleic Acids Res.*, **52**, D442–D455.

18. Meszaros,B., Kumar,M., Gibson,T.J., Uyar,B. and Dosztanyi,Z. (2017) Degrons in cancer. *Sci. Signal.*, **10**, eaak9982.

19. Hou,C., Li,Y., Wang,M., Wu,H. and Li,T. (2022) Systematic prediction of degrons and E3 ubiquitin ligase binding via deep learning. *BMC Biol.*, **20**, 162.

20. Tokheim,C., Wang,X., Timms,R.T., Zhang,B., Mena,E.L., Wang,B., Chen,C., Ge,J., Chu,J., Zhang,W., *et al.* (2021) Systematic characterization of mutations altering protein degradation in human cancers. *Mol. Cell*, **81**, 1292–1308.

21. Xu,H., Hu,R. and Zhao,Z. (2023) DegronMD: leveraging evolutionary and structural features for deciphering protein-targeted degradation, mutations, and drug response to degrons. *Mol. Biol. Evol.*, **40**, msad253.

22. UniProt,C. (2023) UniProt: the Universal Protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.

23. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A, *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

24. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.

25. Zhao,B., Li,L., Tumaneng,K., Wang,C.Y. and Guan,K.L. (2010) A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCF(beta-TRCP). *Genes Dev.*, **24**, 72–85.

26. Oh,J.H., Hyun,J.Y. and Varshavsky,A. (2017) Control of Hsp90 chaperone and its clients by N-terminal acetylation and the N-end rule pathway. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E4370–E4379.

27. Zhang,Z., Sie,B., Chang,A., Leng,Y., Nardone,C., Timms,R.T. and Elledge,S.J. (2023) Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping. *Mol. Cell*, **83**, 3377–3392.

28. Ng,C., Jackson,R.A., Buschdorf,J.P., Sun,Q., Guy,G.R. and Sivaraman,J. (2008) Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-cbl-TKB domain substrates. *EMBO J.*, **27**, 804–816.

29. Yan,X., Wang,X., Li,Y., Zhou,M., Li,Y., Song,L., Mi,W., Min,J. and Dong,C. (2021) Molecular basis for ubiquitin ligase CRL2(FEM1C)-mediated recognition of C-degron. *Nat. Chem. Biol.*, **17**, 263–271.

30. Chen,X., Liao,S., Makaros,Y., Guo,Q., Zhu,Z., Krizelman,R., Dahan,K., Tu,X., Yao,X., Koren,I., *et al.* (2021) Molecular basis for arginine C-terminal degron recognition by Cul2(FEM1) E3 ligase. *Nat. Chem. Biol.*, **17**, 254–262.

31. Tournier,B., Sanchez-Ballesta,M.T., Jones,B., Pesquet,E., Regad,F., Latche,A., Pech,J.C. and Bouzayen,M. (2003) New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett.*, **550**, 149–154.

32. Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.

33. Huerta-Cepas,J., Szklarczyk,D., Heller,D., Hernandez-Plaza,A., Forslund,S.K., Cook,H., Mende,D.R., Letunic,I., Rattei,T., Jensen,L.J., *et al.* (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.*, **47**, D309–D314.

34. Huang,H., Arighi,C.N., Ross,K.E., Ren,J., Li,G., Chen,S.C., Wang,Q., Cowart,J., Vijay-Shanker,K. and Wu,C.H. (2018) iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.*, **46**, D542–D550.

35. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.

36. Willems,P., Horne,A., Van Parys,T., Goormachtig,S., De Smet,I., Botzki,A., Van Breusegem,F. and Gevaert,K. (2019) The Plant PTM Viewer, a central resource for exploring plant protein modifications. *Plant J.*, **99**, 752–762.

37. Xu,H., Zhou,J., Lin,S., Deng,W., Zhang,Y. and Xue,Y. (2017) PLMD: an updated data resource of protein lysine modifications. *J. Genet. Genomics*, **44**, 243–250.

38. Sondka,Z., Dhir,N.B., Carvalho-Silva,D., Jupe,S.,,,Madhumita, McLaren,K., Starkey,M., Ward,S., Wilding,J., Ahmed,M., *et al.* (2024) COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res.*, **52**, D1210–D1217.

39. Oughtred,R., Rust,J., Chang,C., Breitkreutz,B.J., Stark,C., Willems,A., Boucher,L., Leung,G., Kolas,N., Zhang,F., *et al.* (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.*, **30**, 187–200.

40. Del Toro,N., Shrivastava,A., Ragueneau,E., Meldal,B., Combe,C., Barrera,E., Perfetto,L., How,K., Ratan,P., Shirodkar,G., *et al.* (2022) The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.*, **50**, D648–D653.

41. Li,Z., Chen,S., Jhong,J.H., Pang,Y., Huang,K.Y., Li,S. and Lee,T.Y. (2021) UbiNet 2.0: a verified, classified, annotated and updated

database of E3 ubiquitin ligase-substrate interactions. *Database (Oxford)*, **2021**, baab010.

42. Carbon,S., Ireland,A., Mungall,C.J., Shu,S., Marshall,B., Lewis,S., Ami,G.O.H. and Web Presence Working,G. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.

43. Medvar,B., Raghuram,V., Pisitkun,T., Sarkar,A. and Knepper,M.A. (2016) Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation. *Physiol. Genomics*, **48**, 502–512.

44. Rawlings,N.D., Barrett,A.J., Thomas,P.D., Huang,X., Bateman,A. and Finn,R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.*, **46**, D624–D632.

45. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

46. Joosten,R.P., Beek,T.A., Krieger,E., Hekkelman,M.L., Hooft,R.W., Schneider,R., Sander,C. and Vriend,G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **39**, D411–D419.

47. Rost,B. and Sander,C. (1994) Conservation and prediction of solvent accessibility in protein families. *Proteins*, **20**, 216–226.

48. Baek,M., DiMaio,F., Anishchenko,I., Dauparas,J., Ovchinnikov,S., Lee,G.R., Wang,J., Cong,Q., Kinch,L.N., Schaeffer,R.D., *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–876.

49. Erdos,G., Pajkos,M. and Dosztanyi,Z. (2021) IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.*, **49**, W297–W303.

50. Goloborodko,A.A., Levitsky,L.I., Ivanov,M.V. and Gorshkov,M.V. (2013) Pyteomics–a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass. Spectrom.*, **24**, 301–304.

51. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S.E., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker,J.M. (ed). *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp. 571–607.

52. Prokhorenkova,L., Gusev,G., Vorobev,A., Dorogush,A.V. and Gulin,A. (2018) In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Montréal, Canada, pp. 6639–6649.

53. Akiba,T., Sano,S., Yanase,T., Ohta,T. and Koyama,M. (2019) In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, Anchorage, AK, USA, pp. 2623–2631.

54. Pedregosa,F., Varoquaux,G., Gramfort,A., Michel,V., Thirion,B., Grisel,O., Blondel,M., Prettenhofer,P., Weiss,R., Dubourg,V., *et al.* (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

55. Paladin,L., Schaeffer,M., Gaudet,P., Zahn-Zabal,M., Michel,P.A., Piovesan,D., Tosatto,S.C.E. and Bairoch,A. (2020) The feature-viewer: a visualization tool for positional annotations on a sequence. *Bioinformatics*, **36**, 3244–3245.

56. Bevilacqua,M., Paladin,L., Tosatto,S.C.E. and Piovesan,D. (2022) ProSeqViewer: an interactive, responsive and efficient TypeScript library for visualization of sequences and alignments in web applications. *Bioinformatics*, **38**, 1129–1130.

57. Tareen,A. and Kinney,J.B. (2020) Logomaker: beautiful sequence logos in Python. *Bioinformatics*, **36**, 2272–2274.

58. Szczesny,R.J., Kowalska,K., Klosowska-Kosicka,K., Chlebowski,A., Owczarek,E.P., Warkocki,Z., Kulinski,T.M., Adamska,D., Affek,K., Jedroszkowiak,A., *et al.* (2018) Versatile approach for functional analysis of human proteins and efficient stable cell line generation using FLP-mediated recombination system. *PLoS One*, **13**, e0194887.

59. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

60. Kats,I., Khmelinskii,A., Kschonsak,M., Huber,F., Kniess,R.A., Bartosik,A. and Knop,M. (2018) Mapping degradation signals and pathways in a eukaryotic N-terminome. *Mol. Cell*, **70**, 488–501.

61. Hickey,C.M., Breckel,C., Zhang,M., Theune,W.C. and Hochstrasser,M. (2021) Protein quality control degron-containing substrates are differentially targeted in the cytoplasm and nucleus by ubiquitin ligases. *Genetics*, **217**, 1–19.

62. Holt,L.J. (2012) Regulatory modules: coupling protein stability to phopshoregulation during cell division. *FEBS Lett.*, **586**, 2773–2777.

63. Eldeeb,M.A., Ragheb,M.A., Soliman,M.H. and Fahlman,R.P. (2022) Regulation of neurodegeneration-associated protein fragments by the N-degron pathways. *Neurotox. Res.*, **40**, 298–318.

64. Cui,Y., He,S., Xing,C., Lu,K., Wang,J., Xing,G., Meng,A., Jia,S., He,F. and Zhang,L. (2011) SCFFBXL(1)(5) regulates BMP signalling by directing the degradation of HECT-type ubiquitin ligase Smurf1. *EMBO J.*, **30**, 2675–2689.

65. Szulc,N.A., Piechota,M., Biriczová,L., Thapa,P. and Pokrzywa,W. (2023) Lysine deserts and cullin-RING ligase receptors: Navigating untrodden paths in proteostasis. *iScience*, **26**, 108344.

66. Floro,J., Dai,A., Metzger,A., Mora-Martin,A., Ganem,N.J., Cifuentes,D., Wu,C.S., Dalal,J., Lyons,S.M., Labadorf,A., *et al.* (2021) SDE2 is an essential gene required for ribosome biogenesis and the regulation of alternative splicing. *Nucleic Acids Res.*, **49**, 9424–9443.

67. Rageul,J., Park,J.J., Jo,U., Weinheimer,A.S., Vu,T.T.M. and Kim,H. (2019) Conditional degradation of SDE2 by the Arg/N-end rule pathway regulates stress response at replication forks. *Nucleic Acids Res.*, **47**, 3996–4010.

68. Thakran,P., Pandit,P.A., Datta,S., Kolathur,K.K., Pleiss,J.A. and Mishra,S.K. (2018) Sde2 is an intron-specific pre-mRNA splicing regulator activated by ubiquitin-like processing. *EMBO J.*, **37**, 89–101.

69. Aderinwale,T., Bharadwaj,V., Christoffer,C., Terashi,G., Zhang,Z., Jahandideh,R., Kagaya,Y. and Kihara,D. (2022) Real-time structure search and structure classification for AlphaFold protein models. *Commun. Biol.*, **5**, 316.