Briefings in Bioinformatics, 2024, **25(6)**, bbae519
https://doi.org/10.1093/bib/bbae519
Problem Solving Protocol

MetaDegron: multimodal feature-integrated protein language model for predicting E3 ligase targeted degrons

Mengqiu Zheng[1,‡], Shaofeng Lin[2,3,‡], Kunqi Chen[2,3], Ruifeng Hu[4], Liming Wang[5,*], Zhongming Zhao[4,6,*], Haodong Xu[1,4,*]

[1]Department of Orthopaedics, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China
[2]Key Laboratory of Gastrointestinal Cancer (Fujian Medical University), Ministry of Education, School of Basic Medical Sciences, Fuzhou 350004, China
[3]Fujian Key Laboratory of Tumor Microbiology, School of Basic Medical Sciences, Fujian Medical University, Fuzhou 350004, China
[4]Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, United States
[5]School of Biomedical Science, Hunan University, Changsha, Hunan, China
[6]MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, United States

*Corresponding authors. Liming Wang, School of Biomedical Science, Hunan University, Changsha, Hunan 410082, China. E-mail:wangliming@hnu.edu.cn; Zhongming Zhao, Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, United States. E-mail:zhongming.zhao@uth.tmc.edu; Haodong Xu, Department of Orthopaedics, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China. E-mail: xuhaodong@csu.edu.cn

‡Mengqiu Zheng and Shaofeng Lin contributed equally.

Abstract

Protein degradation through the ubiquitin proteasome system at the spatial and temporal regulation is essential for many cellular processes. E3 ligases and degradation signals (degrons), the sequences they recognize in the target proteins, are key parts of the ubiquitin-mediated proteolysis, and their interactions determine the degradation specificity and maintain cellular homeostasis. To date, only a limited number of targeted degron instances have been identified, and their properties are not yet fully characterized. To tackle on this challenge, here we develop a novel deep-learning framework, namely MetaDegron, for predicting E3 ligase targeted degron by integrating the protein language model and comprehensive featurization strategies. Through extensive evaluations using benchmark datasets and comparison with existing method, such as Degpred, we demonstrate the superior performance of MetaDegron. Among functional features, MetaDegron allows batch prediction of targeted degrons of 21 E3 ligases, and provides functional annotations and visualization of multiple degron-related structural and physicochemical features. MetaDegron is freely available at http://modinfor.com/MetaDegron/. We anticipate that MetaDegron will serve as a useful tool for the clinical and translational community to elucidate the mechanisms of regulation of protein homeostasis, cancer research, and drug development.

Keywords: targeted protein degradation; ubiquitin-proteasome system; E3 ligase; degrons; deep-learning; web server

Introduction

Cells employ protein degradation to eliminate damaged, abnormal, misfolded, and other unnecessary proteins [1, 2]. In eukaryotic cells, protein degradation primarily occurs through the ubiquitin-proteasome system (UPS) [3]. This process is not only vital for maintaining protein homeostasis but also essential for ensuring the proper functioning of cellular processes such as cell cycle progression, signal transduction, differentiation, and growth [4, 5]. Dysfunction in protein degradation can lead to various diseases, including malignant tumors and neurodegenerative disorders [6–8]. Ubiquitin (Ub), a highly conserved protein composed of 76 amino acids with a molecular weight of 8 kDa, plays a central role in this process. The coordinated action of ubiquitin ligases, including the E1: Ub-activating enzyme, E2: Ub-conjugating enzyme, and E3: Ub-ligase, triggers a cascade reaction that attaches ubiquitin to the protein designated for degradation, forming a ubiquitinated protein complex. In the initial step, E1 activates Ub in the presence of adenosine triphosphate and binds the C-terminus of Ub to the active site of E1 (step 1). Subsequently, the activated Ub binds to the cysteine residue in the active site of E2, transferring the activated Ub to E2 (step 2). With the catalytic assistance of E3, the carboxyl-terminal glycine of ubiquitin conjugates to the amino group of the substrate protein residue, typically lysine, resulting in the ubiquitination of the substrate protein (step 3) [9–11].

Degradation signals (degrons) are short linear amino acid motifs, located on target protein substrates [12–14]. When a protein receives a degradation signal, the degron becomes exposed and recognized, after which it is bound by E3 ubiquitin ligase [15, 16]. This binding facilitates the entry of ubiquitinated protein substrates into the enzyme network centered on the 26S proteasome for degradation [17]. The interaction between E3 ligase and the degron is highly specific, playing a pivotal role in determining degradation specificity and maintaining cellular homeostasis [18–20]. An example of this mechanism is the degradation-specific binding of MDM2 (E3) to the degron in the tumor suppressor protein p53, which facilitates the targeted

Received: July 10, 2024. **Revised:** September 10, 2024. **Accepted:** October 2, 2024

degradation of p53 [14, 21, 22]. Under normal conditions, despite continuous production, p53 is kept at low levels in the cytoplasm due to its ongoing degradation by the UPS. However, when cells are stimulated by external signals, changes in the degron sequence or structure prevents binding to MDM2, resulting in abnormal accumulation of p53 within the cell. This abnormal accumulation then triggers a series of cellular cytotoxic effects, such as cell cycle arrest and apoptosis [6, 23]. The identification of E3-degron interactions is fundamental to understanding the dynamic regulation of proteins.

Recently, the development of high-throughput experimental techniques and proteomics technologies has expanded our understanding of E3s and degrons within the UPS, thereby accelerating the development of targeted protein degradation-based drug therapies and bringing hope to numerous patients [13, 23–29]. However, the identification of degrons remains challenging due to the undetermined substrates for most E3s. The nature of degron and publicly available data sets make it possible to develop computational methods to identify it using pattern recognition and machine learning techniques, which facilitates the development of a number of bioinformatic resources/tools available for degron identification [30–32]. The APC/C degron repository provides valuable insights into the determinants of APC/C degron sequences [33], encompassing information regarding to disordered regions and post-translational modifications. Complementarily, the Eukaryotic Linear Motif (ELM) resource catalogues numerous degron motifs present in proteins [34]. Specifically, DegronMD serves as a novel resource tailored for the comprehensive exploration of degrons, encompassing associated functional aberrations and responses to pharmacological treatments [35]. This resource contains 23 distinct internal degrons within the human proteome. Augmenting these bioinformatic repositories, deep learning models such as Degpred [30] and deepDegron [28] have emerged for the prediction of degrons and the assessment of their perturbations by mutations, respectively. Notably, deepDegron predicts the likelihood of a protein sequence harboring N- or C-terminal degrons, mandating specific mutation input files. Conversely, Degpred leverages a BERT-based architecture to primarily predict internal degrons based solely on sequence information. Additionally, DEGRONOPEDIA [31] has been introduced as a novel web server dedicated to the identification and analysis of degron motifs within proteins, enabling the prediction of potential N-/C-degrons subsequent to proteolytic events (summarized in Supplementary Table S1).

Although the early methods for screening the binding of E3 targeted degrons showed promising, the majority of the models were developed by solely utilizing motif matching or sequence-based machine learning models, which prevents them from learning the whole complex features of degrons, especially the structural properties. In addition, none of the previous methods build an online responsive platform to provide timely and customized predictions for E3 targeted degron, restricting its capability for degron identifications. In this study, we present the MetaDegron, a novel bioinformatics tool accompanied by a user-friendly web service, designed to predict E3 targeted degron. MetaDegron incorporates comprehensive featurization strategies and leverages the protein language model to identify novel degron instances, which was trained on a curated dataset of 300 degron instances. Moreover, extensive evaluation and comparative analysis demonstrate the superior performance of MetaDegron. Functionally, MetaDegron offers the convenience of batch prediction for targeted degrons associated with 21 E3 ligases. Furthermore, the web service provides functional annotations and visualization tools for a range of degron-related structural and physicochemical features. MetaDegron can be accessed freely at http://modinfor.com/MetaDegron/ and https://github.com/BioDataStudy/MetaDegron, enabling broad accessibility to the research community and facilitating exploration within the fields of biological mechanisms, protein degradation implications, and degron-centric drug development.

Methods

Data preparation

We collected and processed a set of human degron motifs, which are E3 binding consensus patterns, from the ELM database [36], and over 300 degron instances from a number of previous studies (Supplementary Table S2) [14, 30, 32, 37]. For instance, β-TrCP2 (E3, known as FBXW11) recognizes DSGxxS consensus degron motif, where x denotes any one of the 20 amino acids. In order to improve the characterization of degrons, we constructed a comprehensive background dataset comprising a considerable number of randomly selected peptides with the same length as the degron instances (Supplementary Table S3). This strategy enabled us to perform comparative analysis based on peptide sequences of similar length, thus providing a suitable control to evaluate the specificity of degron sequences. To further investigate the structural and physicochemical properties associated with the curated degrons, multiple analyses were performed [38–45]. We calculated 10 features for all motif instances or random peptides. Specifically, the determination of residue-specific flexibility utilized the DynaMine software [40] employing default parameters. Residue-specific solvent accessibility and secondary structures, including coiled coil and α-helix, were computed using the Spider2 tool [39]. Protein disorder was assessed through the utilization of the IUPred software [38]. The anchoring score of each degron was evaluated employing the ANCHOR program [46]. Multiple sequence alignments (MSA) of orthologous proteins were acquired utilizing the Gopher tool from Bioware [47] to calculate sequence conservation. Information pertaining to protein domains was retrieved from the Pfam database [41]. Moreover, we evaluated the enrichment of important post-translational modifications (PTMs, phosphorylation and ubiquitination) within and around degrons. The experimentally verified PTMs information was downloaded from our constructed Eukaryotic Phosphorylation Sites Database [43] and Protein Lysine Modifications Database resources [42], respectively. By comprehensively analysing and elucidating the properties of these degrons, we can gain valuable insights into their sequence motifs, structural features, and biochemical properties. Such insights not only enhance our understanding of protein degradation but also facilitate the development of more accurate prediction models. To facilitate the development and evaluation of our models, the constructed dataset was partitioned into a training dataset, which represented 90% of the total data, and an independent dataset, which accounted for the remaining 10% of the data.

Model architecture

Based on curated degrons and random peptides, we employed a bootstrapping strategy to train 10 eXtreme Gradient Boosting (XGBoost) classifiers [48], leveraging multiple distinguishing features between these two groups. Exhaustively exploring numerous parameter combinations for each classifier, we then selected the optimal parameters through multiple rounds of cross-validation (CV) assessments. The average prediction scores of these 10 models were employed to establish the ultimate probability of degron occurrence. To ensure that the model can make predictions when the protein features are not matched, we extended its functionality by incorporating a deep neural network (DNN).

The DNN architecture employed by MetaDegron comprises two distinct components, as illustrated in Fig. 3. The initial component utilizes context-sensitive embeddings of amino acids generated by Embeddings from Language Models (ELMo) [49], known for its efficacy in various protein sequence prediction tasks. ELMo is a contextual word embedding technique originally developed for natural language processing tasks. Here, we used SeqVec, a protein-specific adaptation of ELMo, leverages bidirectional Long Short-Term Memory (BLSTM) networks to generate context-sensitive embeddings for amino acid sequences. By modeling protein sequences, SeqVec effectively captured the biophysical properties of the language of life from unlabeled big data (UniRef50), which allows SeqVec to generate embeddings that encode the contextual dependencies between amino acids across the entire sequence. Each amino acid in a given protein sequence is encoded into a 1024-dimensional vector, encapsulating chemical, physical, and structural information. These embeddings are then used as inputs for a BLSTM layer and a convolution-pooling layer, which are sequentially connected to a dense layer for feature integration.

By employing convolutional layers followed by max-pooling operations, the convolutional neural network (CNN) component of our model can effectively extract hierarchical features representing short-range interactions between amino acid residues. These features capture local sequence motifs that may be indicative of E3 ligase targeting signals or degron recognition motifs. The second component of the DNN architecture involves encoding each amino acid into an adjustable-length vector, which is context-insensitive and undergoes joint training with the rest of the DNN. This embedding layer is connected to a BLSTM layer followed by another dense layer. Here, we leverage word embedding techniques, inspired by natural language processing, to transform each amino acid residue within a peptide segment into a dense, continuous vector representation. Similarly, in the context of peptide sequences, we can exploit the inherent similarities and relationships between amino acids to learn meaningful representations that capture their contextual dependencies. By integrating a BLSTM network with the word embedding model, we enable the model to capture both forward and backward sequential dependencies within peptide sequences. Subsequently, the two dense layers are further connected to an additional pair of dense layers, ultimately culminating in an output layer with two nodes representing the degron and random peptide classes.

Model evaluation

In this study, we adopt a rigorous validation framework that encompasses both CV techniques and independent test sets to comprehensively evaluate the performance of MetaDegron. Firstly, we employed five-fold CV, a widely used technique in machine learning, to assess the model's performance on the training dataset. This involves partitioning the dataset into k subsets, training the model on k-1 subsets, and evaluating its performance on the remaining subset. This process is repeated k times, with each subset serving as the validation set once. By averaging the performance metrics across multiple iterations, we obtain a more reliable estimate of the model's predictive performance and robustness to variations in the training data. Specifically, For the MetaDegron model, receiver operating characteristic (ROC) curves were constructed for each subset and five ROC curves were generated, and subsequently, a mean ROC curve was calculated, ensuring that each of the five models carried equal weight. In order to assess the model's generalizability, ROC curves were plotted, and the corresponding area under the ROC curve (AUC) values were determined using an independent dataset. In general, the AUC values range from 0 to 1, while a higher AUC value indicates a higher accuracy of a predictive model. Moreover, additional metrics, e.g., the area under the precision-recall curve (AUPRC), accuracy, recall and F1-score, were also calculated for a comprehensive comparative analysis. We also compared the performance of the MetaDegron with previous method using an additional dataset comprising experimentally validated substrates of β-TrCP2 (Supplementary Table S4). To avoid the possible bias caused by a random sampling from the background dataset, the random sampling was performed 10 times, and the same number of random peptides with the same length was extracted each time. We calculated the average AUC, AUPRC, accuracy, recall, and F1-score.

Website implementation

The MetaDegron web server was developed following the standard Model-View-Controller framework, a prevalent approach in contemporary web application design [50]. This framework was structured into three primary logical components, namely 'Prediction', 'Results', and 'Controller', which collectively constitute the MetaDegron system. On the backend, the system integrates two well-optimized models, specifically XGBoost and deep learning models denoted as MetaDegron-X and MetaDegron-D, tailored for the prediction of E3 ligase-targeted degrons. The 'Controller' module serves a pivotal role by validating the input data's format, facilitating the transfer of data from the frontend interfaces to the backend, orchestrating the execution of predictive models, and ultimately delivering the results to the 'Results' page. The 'Prediction' component, positioned as the frontend interface, enables user interactions with the system. To ensure a responsive server, both the 'Prediction' and 'Results' interfaces were constructed using an amalgamation of HTML5, CSS (utilizing Bootstrap3), and JavaScript. jQuery, a JavaScript library, was employed to leverage Ajax technology for seamless communication with the 'Controller' module. Additionally, PHP was employed as a complementary tool for the presentation of results. In the 'Results' page, the MetaDegron provides the functional annotations and visualizations of degron-related features and source protein. For the degron, the properties calculated by MetaDegron are provided for users. For the source protein, the base information is obtained to display from the UniProt database. The interacting proteins of source protein, including E3s and deubiquitinating enzymes (DUBs) are provided with a tabular list and an interactive network based on the Cytoscape.js [51]. For both degron and source protein, the structural and MSA information are visualized by 3Dmol.js [52] and ProViz tool [53], respectively.

Results

Overview of the MetaDegron framework

MetaDegron is a novel tool specifically designed for precise degron prediction using machine learning techniques (Fig. 1). It comprises two distinct models: MetaDegron-D and MetaDegron-X, each employing different methodologies for degron prediction. MetaDegron-D operates by extracting features solely from the protein sequence. The input to this model is peptide segments

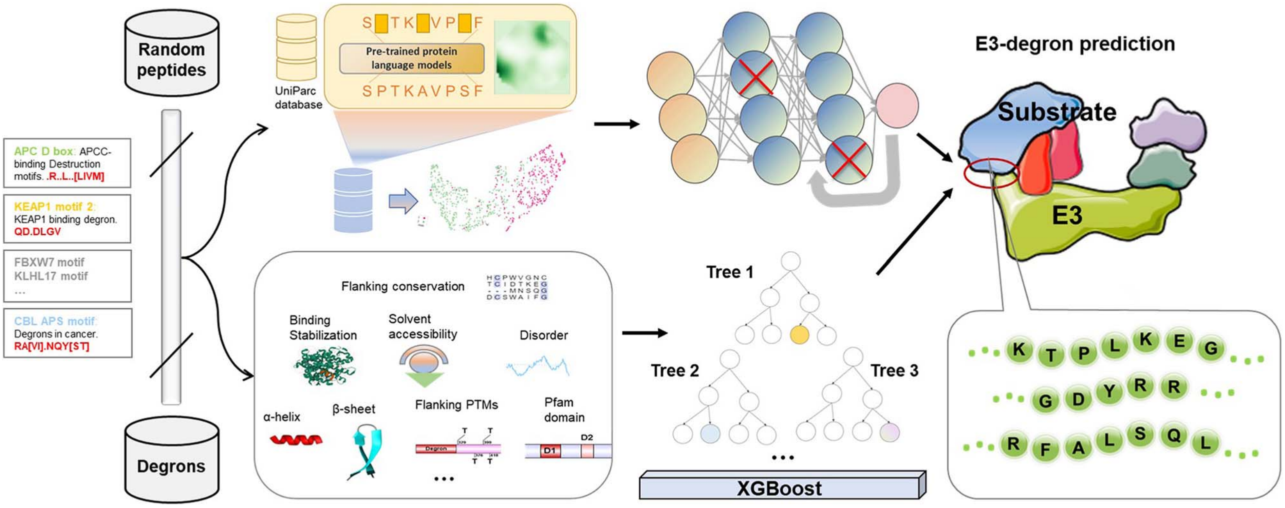


Figure 1. The overall framework of MetaDegron.

that are centered around the degron region. We employ a pre-trained protein model to represent each degron, which generates context-sensitive embeddings in a 1024-dimensional space. These embeddings undergo further processing using various feature extraction networks such as convolution-pooling or BLSTM layers. In addition, another word embedding model integrating BLSTM network was utilized to convert the peptide segment into a 27-dimensional vector for pattern extraction. These two components are combined through pairing of these fully connected networks, resulting in the final degron prediction. On the other hand, MetaDegron-X takes into consideration multiple distinctive features of a degron itself. This model employs input features that encompass various aspects of the degron, including sequence, evolution, and structure. To achieve this, an XGBoost classifier is employed for final prediction. Overall, MetaDegron possesses the capability to predict targeted degrons of 21 E3 ligases in a batch manner, and provides functional annotations and visualization of multiple degron-related structural and physicochemical features. The comprehensive functionality of MetaDegron allows researchers to gain crucial insights into the functional aspects of degrons and their relation to a wide range of protein characteristics.

Characterization and prediction of degron with structure characteristics

Targeted degrons play pivotal roles in regulating protein stability and turnover within the cell, influencing various cellular processes. Understanding the structure characteristics of degrons is crucial for predicting their degradation potential and unraveling their functional implications. To characterize the structure features of degrons, we analyzed curated, experimentally validated degrons by comparing them to a background dataset. We employed various structural bioinformatics algorithms and tools to identify common structural properties within degrons [36, 38, 39, 41–43, 45]. Remarkably, the known degrons exhibited a higher degree of solvent accessibility and binding stability compared to the random peptides (Fig. 2A and B), suggesting their importance in recognition by degradative enzymes. Furthermore, degrons were found to be preferentially located in protein disordered regions (Fig. 2C), highlighting their distinctive localization patterns. Additionally, the analysis revealed a specific preference of degrons for coiled coil regions rather than α-helix regions (Fig. 2D and E). It was also observed that degrons tend to occur in lower flexibility regions (Fig. 2F). These findings provided valuable insights into the structural characteristics of degrons and indicate potential determinants for degron recognition and degradation. Subsequently, the XGBoost classifier (called MetaDegron-X) was constructed using these discerning features for E3 targeted degron. The performance of MetaDegron-X, as assessed by the AUC values, was promising. Specifically, the AUC values ranged from 0.83 to 0.89 in a five-fold CV, with an average AUC value of 0.87 (Fig. 2G). Furthermore, validation of the developed MetaDegron-X was carried out on an independent testing dataset. The performance of MetaDegron-X was superior, as denoted by the AUC value of 0.86 (Fig. 2H). These findings collectively demonstrated the high accuracy and robust performance achieved by MetaDegron-X.

In addition, we conducted an elimination study to assess the association between each feature and the prediction results. Specifically, we systematically removed one feature at a time from the input data and evaluated the impact on the predictive performance of our model. By quantifying the change in AUC value through five-fold CV, we revealed the relative importance of each feature in predicting E3 ligase targeted degrons (Fig. 2I). From the results, we found all features contribute to the construction of the model, i.e. the performance of the models decreased to different degrees after removing the specific features. Overall, those features, such as disorder and the number of PTMs within the degron region, represented the most important features. This is consistent with the facts that degrons are preferentially located in disordered regions and regulated by PTMs that control the interaction with E3s in response to environmental and cellular cues.

Enhancing the MetaDegron system using deep learning technology

The inability to match protein features on certain occasions may pose a challenge for feature-driven MetaDegron-X to achieve prediction. To address this issue and enhance the prediction capabilities of the model, we have developed an extended version called MetaDegron-D. By incorporating a deep learning framework (see 'Methods' part), MetaDegron-D was capable of solely operating on protein sequences. This novel approach utilized a hybrid architecture comprising cutting-edge deep learning

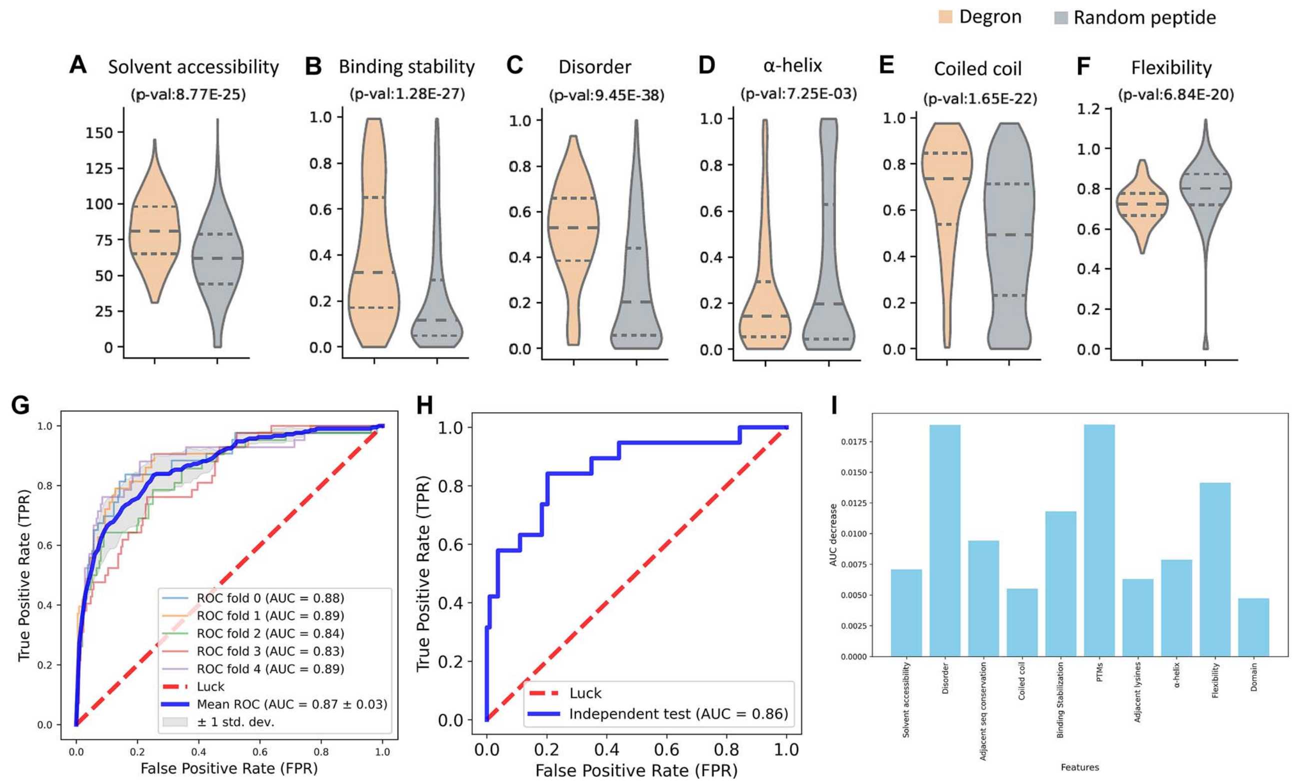


Figure 2. Performance of MetaDegron based on protein features. (A-F) the statistics of multiple characteristic features for the known degron instances and random peptides, including solvent accessibility (A), binding stability (B), disorder (C), α-helix (D), coiled coil (E), and flexibility (F). (G-H) ROC curves for MetaDegron-X in different five-fold CV (G) and independent testing dataset (H). (I) Elimination study to assess the association between each feature and the prediction results.

networks (Fig. 3A), such as protein language models, word embeddings, convolution, and BLSTM, as thoroughly detailed in the methodology section. This deep learning framework allows MetaDegron-D to leverage the full potential of these advanced networks and their ability to extract high-level features from protein sequences. The performance evaluation of MetaDegron-D demonstrated its great predictive capabilities. Through a five-fold CV approach, we obtained an average AUC value of 0.90. Furthermore, the AUC values ranged from 0.89 to 0.92, indicating consistent and reliable performance (Fig. 3B). Additionally, when tested with an independent dataset, MetaDegron-D achieved an improved AUC value of 0.90 (Fig. 3C). These results suggested the robustness and accuracy of MetaDegron-D in predicting protein features solely from the sequence information.

Moreover, we assessed the performance of the MetaDegron using an additional dataset comprising experimentally validated substrates of β-TrCP2, as detailed in Supplementary Table S4. We first extracted all reported substrates, and matched and filtered those with an instance of β-TrCP2 degron within their protein sequence. A subset consisting of 20 substrates exhibiting the presence of the β-TrCP2 degron was designated as the positive dataset. We also constructed a background dataset comprising over 700 randomly selected peptides with the same length as the degron instances (Supplementary Table S4). Using this new benchmark dataset, we compared MetaDegron to Degpred and found that MetaDegron could achieve better AUROC value (Fig. 3D). The AUC values were computed as 0.9705 (ranged from 0.9275 to 0.9950, MetaDegron-D), 0.9670 (ranged from 0.9425 to 0.9875, MetaDegron-X), and 0.9540 (ranged from 0.9300 to 0.9700, Degpred), respectively. We also computed the additional metrics, e.g. AUPRC, accuracy, recall, and F1-score, for our models and Degpred. MetaDegron still outperformed Degpred in term of these metrics (Supplementary Table S5). Through extensive evaluation on benchmark datasets and comparison with Degpred, we demonstrated the superior performance of MetaDegron.

To further explore the capabilities of the MetaDegron framework, we utilized the visualization method described by Becht et al. [54] to compare the features of degrons and random peptides across each network layer (Fig. 3E-I). As expected, the feature representations of the input layer for both degrons and random peptides exhibited significant overlap and mixing (Fig. 3E). However, as the framework underwent training, a clear distinction between degrons and random peptides emerged, resulting in more separated clusters within the feature space (Fig. 3I). This observation emphasizes the effectiveness of the MetaDegron-D model in identifying efficient features and differentiating between degrons and random peptides.

The usage of MetaDegron

MetaDegron serves as a useful tool for predicting targeted degrons of 21 E3 ligases, offering researchers possible candidates for studying protein degradation pathways and identifying potential therapeutic targets. The multimodal feature integration approach enables MetaDegron to capture diverse aspects of degron recognition, including amino acid composition, physicochemical properties, evolutionary conservation, and contextual dependencies, thereby enhancing its capabilities for advancing research in the field of protein degradation and ubiquitin-mediated proteolysis. The webserver of MetaDegron was designed and constructed with a modular and user-friendly manner (Fig. 4). Three major modules, including ‘Run’, ‘Results’ and ‘Tutorial’, are the kernel of

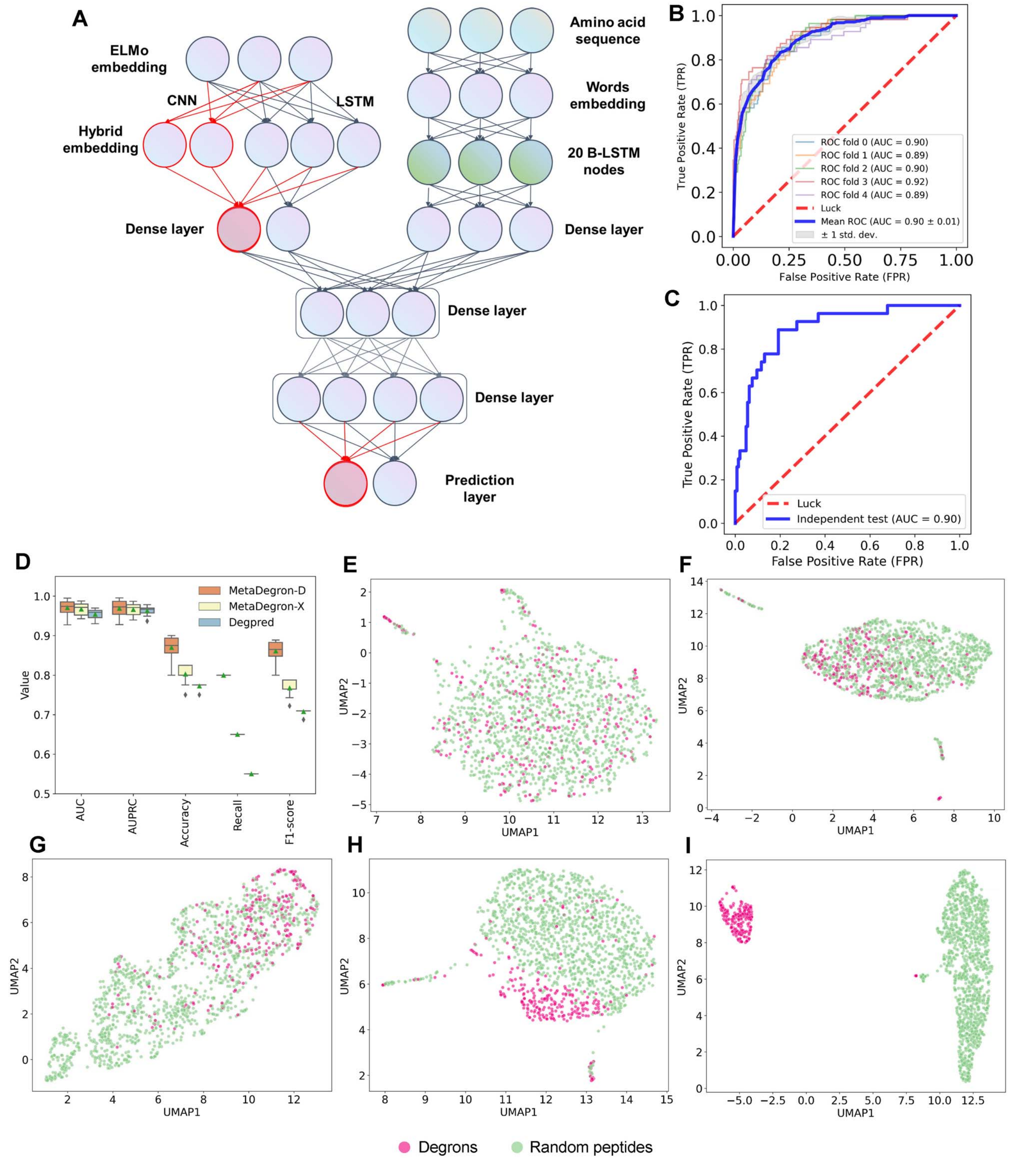
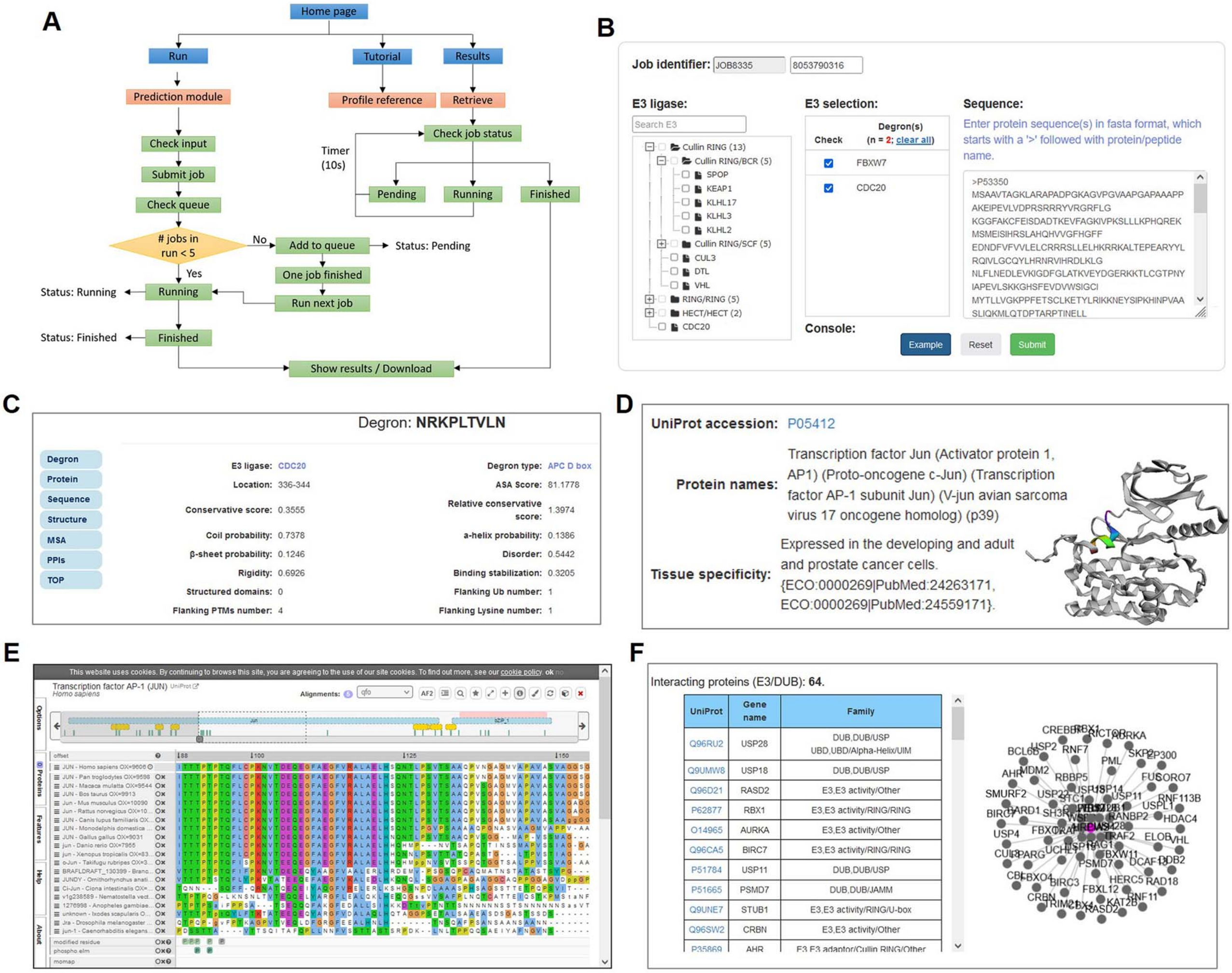


Figure 3. Implementation and performance of MetaDegron based on deep learning framework. (A) The hybrid deep learning architecture of MetaDegron. (B-C) ROC curves for MetaDegron-D in different five-fold CV (B) and independent testing dataset (C). (D) Comparison of MetaDegron with Degpred on the additional independent dataset. (E-I) feature representation of the known degrons and random peptides using the UMAP method in each network layer, including input layer (E), dense layer (context-sensitive) (F), dense layer (context-insensitive) (G), joint features layer (H), dense layer (before output) (I).

MetaDegron online server (Fig. 4A). The 'Run' module sequentially controls the execution of submitted jobs, including the input checking, job submitting, job running, and task terminates. Meanwhile, the 'Results' module records the submission jobs, monitors the status of jobs, and immediately shows the prediction results. The clickable and searchable hierarchical classification tree of E3s is loaded for the selection of single or multiple E3 ligases (Fig. 4B). Then, one or more protein sequences in FASTA format can be submitted. After finishing the submitted job, the prediction results will be visualized with specific information, including the 'Entry', 'E3 ligase', 'Degron instance', 'Degron type', 'Start', 'End', and 'Score'. It displays the detailed information for degron



Job identifier: JOB8335 8053790316

E3 ligase:

Search E3

- □ Cullin RING (13)
  - □ Cullin RING/BCR (5)
    - □ SPOP
    - □ KEAP1
    - □ KLHL17
    - □ KLHL3
    - □ KLHL2
  - □ Cullin RING/SCF (5)
  - □ CUL3
  - □ DTL
  - □ VHL
- □ RING/RING (5)
- □ HECT/HECT (2)
- □ CDC20

E3 selection:

Check	Degron(s) (n = 2; clear all)
☑	FBXW7
☑	CDC20

Sequence:

Enter protein sequence(s) in fasta format, which starts with a '>' followed with protein/peptide name.

```
>P53350
MSAAVTAGKLARAPADPGKAGVPGVAAPGAPAAAPP
AKEIPEVLVDPRSRRRYVRGRFLG
KGGFAKCFEISDADTKEVFAGKIVPKSLLLKPHQREK
MSMEISIHRSLAHQHVVGFHGFF
EDNDFVFVVLELCRRRSLLELHKRRKALTEPEARYYL
RQIVLGCQYLHRNRVIHRDLKLG
NLFLNEDLEVKIGDFGLATKVEYDGERKKTLCGTPNY
IAPEVLSKKGHSFEVDVWSIGCI
MYTLLVGKPPFETSCLKETYLRIKKNEYSIPKHINPVAA
SLIQKMLQTDPTARPTINELL
```

Console:

Example Reset Submit

Degron: **NRKPLTVLN**

Degron | Protein | Sequence | Structure | MSA | PPIs | TOP

E3 ligase: CDC20
Degron type: APC D box
Location: 336-344
ASA Score: 81.1778
Conservative score: 0.3555
Relative conservative score: 1.3974
Coil probability: 0.7378
a-helix probability: 0.1386
β-sheet probability: 0.1246
Disorder: 0.5442
Rigidity: 0.6926
Binding stabilization: 0.3205
Structured domains: 0
Flanking Ub number: 1
Flanking PTMs number: 4
Flanking Lysine number: 1

UniProt accession: P05412

Protein names: Transcription factor Jun (Activator protein 1, AP1) (Proto-oncogene c-Jun) (Transcription factor AP-1 subunit Jun) (V-jun avian sarcoma virus 17 oncogene homolog) (p39)

Tissue specificity: Expressed in the developing and adult and prostate cancer cells. {ECO:0000269|PubMed:24263171, ECO:0000269|PubMed:24559171}.

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our cookie policy. ok no

Transcription factor AP-1 (JUN) UniProt
Homo sapiens

Alignments: 5 qfo AF2

Interacting proteins (E3/DUB): **64**.

UniProt	Gene name	Family
Q96RU2	USP28	DUB,DUB/USP UBD,UBD/Alpha-Helix/UIM
Q9UMW8	USP18	DUB,DUB/USP
Q96D21	RASD2	E3,E3 activity/Other
P62877	RBX1	E3,E3 activity/RING/RING
O14965	AURKA	E3,E3 activity/Other
Q96CA5	BIRC7	E3,E3 activity/RING/RING
P51784	USP11	DUB,DUB/USP
P51665	PSMD7	DUB,DUB/JAMM
Q9UNE7	STUB1	E3,E3 activity/RING/U-box
Q96SW2	CRBN	E3,E3 activity/Other
P35869	AHR	E3,E3 adaptor/Cullin RING/Other

Figure 4. The usage of MetaDegron webserver. The general pipeline for MetaDegron server. (B) The example selection of E3s and sequences in the ‘prediction’ page. (C) The feature properties of selected degron. (D-F) The annotations of source protein for selected degron, including the base and structure information (D), the MSA viewer (E), and the interacting annotations (F).

and source protein (Fig. 4C-F). The properties of degron (Fig. 4C) and information of source protein (Fig. 4D) are displayed as well. In addition, the structure of source protein is presented with 3Dmol.js [52], and the degron instance is marked with highlights. Moreover, the MSA of degron instance and source protein are visualized by using the ProViz tool [53] (Fig. 4E), and the interacting E3s or DUBs of source protein are provided in a tabular list and an interactive network based on the Cytoscape.js [51] (Fig. 4F). Taken together, MetaDegron is a user-friendly online tool for the study of targeted protein degradation.

Discussion

Targeted protein degradation represents a highly promising therapeutic modality that is currently gaining considerable attention in the biomedical fields [5, 8, 55–60]. In the ubiquitin proteasome system, the interactions between E3 ligase and the degradation signal (degron) is critical for determining degradation specificity and maintaining cellular homeostasis. In this study, we develop a new approach, called MetaDegron, for predicting E3 targeted degron, based on our careful data collection, curation, and analysis. The built-in models of MetaDegron integrate comprehensive featurization strategies and large protein language model, displaying great performance in both machine learning and deep leaning models. MetaDegron represents a novel framework that integrates state-of-the-art protein language models with a diverse array of featurization techniques to enhance the predictive accuracy of E3 ligase targeted degrons. By leveraging protein language models, MetaDegron can effectively extract high-dimensional embeddings that encode rich contextual information from the input protein sequences. Moreover, MetaDegron utilizes a multi-modal feature integration approach, which combines contextual information, sequence-based features, and structural features to improve prediction performance. This innovative tool first time enables users to perform batch prediction of targeted degrons for 21 E3 ligases, while it also provides functional annotations and visualizations of various degron-related structural and physicochemical features.

Our structural analyses in this study help a deep understanding of the structural characteristics of degrons. Accordingly, we developed a robust computational method, MetaDegron-X, for predicting E3-targeted degrons. The identification of degron structural properties and the development of accurate prediction tools are crucial steps towards unraveling the regulatory mechanisms underlying protein degradation processes and may have implications for drug discovery and biomedical research [15, 61]. Moreover, our extended MetaDegron-D model demonstrates excellent predictive performance, indicating its potential to overcome the challenges posed by mismatched protein features. The incorporation of a deep learning framework enables MetaDegron-D to operate solely on protein sequence information and achieve accurate predictions. The visualization of degrons and random peptides further supports the efficacy of the MetaDegron-D model in distinguishing between these two classes based on their specific features.

The protein homeostasis primarily depends on the protein degradation via UPS, and the aberrant regulations of protein homeostasis can lead to various diseases, including cancer, neurodegenerative disorders, and inflammatory conditions. Predicting E3 ligase targeted degrons holds significant promise for accelerating our understanding of protein degradation pathways and facilitating the discovery of novel therapeutic targets. By identifying specific sequences recognized by E3 ligases for ubiquitination and subsequent degradation, MetaDegron enables the characterization of substrate specificity and regulatory mechanisms of E3 ligases, shedding light on their roles in protein degradation and homeostasis. In addition, predicting E3 ligase targeted degrons can provide critical insights into the aberrant protein turnover mechanisms driving oncogenesis and tumor progression. By elucidating the substrate specificity and regulatory networks of E3 ligases in cancer cells, researchers can identify therapeutic targets for precision oncology interventions. Moreover, MetaDegron's predictive capabilities have implications for drug discovery and development efforts targeting protein degradation pathways. By enabling the identification of candidate substrates for E3 ligases, MetaDegron can facilitate the screening and prioritization of compounds that modulate protein degradation processes. Finally, MetaDegron may aid in the design of targeted therapies aimed at selectively inducing protein degradation of disease-associated proteins, offering a new paradigm for drug development. Overall, predicting E3 ligase targeted degrons has far-reaching implications for basic research, translational studies, cancer research, and drug development. By providing insights into degradation mechanisms, identifying therapeutic targets, and guiding precision medicine approaches, we anticipate that the MetaDegron can serve as a useful tool to identify E3-targeted degrons for further research of protein regulation and drug development.

There are some limitations in this study. First, degrons are short linear motifs specifically recognized by E3s; therefore, the proteins with similar degrons may be recognized by a specific E3 to further off-target degradation. However, the interaction of E3 and degron really provides the potential approach to degrade the drug target. Meanwhile, the off-target effects of the predicted molecule can be reduced in the design of targeted drug, such as the consideration of more molecular features of target protein, the discovery of ligand, and the development the molecular glue. Another potential limitation lies in the biases present in the datasets used for training and evaluation. Experimental studies characterizing degron sequences may exhibit biases towards certain protein families, cellular contexts, or experimental techniques. Consequently, the predictive performance of our model may be influenced by the distribution of data across different classes and may not fully generalize to unseen data or underrepresented classes. Especially, the number of known degron instances, reported E3-degron interactions, and E3-substrate interactions are still limited. In addition, our approach relies on the integration of diverse features, including sequence-based, structural, and contextual features, to characterize degron sequences comprehensively. However, the selection and representation of these features may introduce implicit assumptions about the underlying mechanisms of E3 ligase targeting and degron recognition. While we aim to capture a broad range of sequence-structure relationships, our feature representation may not fully capture all relevant aspects of degron recognition, leading to potential limitations in predictive performance. Also, while our model provides accurate predictions of E3 ligase targeted degrons, the interpretability of these predictions may be limited. It may be challenging to elucidate the underlying biological mechanisms, especially for complex or novel sequences. More experimental validation is necessary to further optimize the reliability of the predictions. Enhancing the interpretability of our model's predictions could improve its utility for guiding experimental validation and hypothesis generation. In terms of performance comparison, both deepDegron and DEGRONOPEDIA were developed for predicting N-/C-degrons, whereas only Degpred supported the prediction of internal degron and E3-degron interactions. Thus, only Degpred tool was used for performance comparison in this study. We expect more tools to be developed in the future, leading to more extensive comparisons.

In future, more useful features and machine learning frameworks will be adopted for the improvement of MetaDegron models. We expect more E3 and their targeted degrons will be discovered, and they will be integrated into our framework to extend the benchmark dataset and improve the performance of MetaDegron. Moreover, the functional prediction of E3s and targeted degrons should be considered by combining the high-throughput omics and computational prediction. We will continuously maintain and improve the server of MetaDegron.

Key Points

- We integrate the protein language model and comprehensive featurization strategies to develop a novel framework, namely MetaDegron for the identification of new degron for targeted protein degradation.
- MetaDegron covers 21 E3 ligases-targeted degron predictions, showing excellent performance through extensive evaluation and comparative analysis.
- MetaDegron implements an online prediction website and provides functional annotations and visualization of multiple degron-related structural and physicochemical features.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

Funding

This work was supported by grants from the Natural Science Foundation of China (Grant No. 32300528 and 32300520), the Science and Technology Innovation Program of Hunan province (Grant No. 2023RC3080), the Excellent Youth Foundation of Hunan Scientific Committee (Grant No. 2024JJ2084) and the Scientific Research Fund of Hunan Provincial Education Department (Grant No. 23B0023), and Scientific Research Foundation for Advanced Talents of Fujian Medical University (Grant No. XRCZX2022015).

Data availability

MetaDegron can be accessed freely at http://modinfor.com/MetaDegron/ and https://github.com/BioDataStudy/MetaDegron.

References

1. Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. *Nature* 2003;**426**:895–9. https://doi.org/10.1038/nature02263.
2. Pohl C, Dikic I. Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science* 2019;**366**:818–22. https://doi.org/10.1126/science.aax3769.
3. Nalepa G, Rolfe M, Harper JW. Drug discovery in the ubiquitin–proteasome system. *Nat Rev Drug Discov* 2006;**5**:596–613. https://doi.org/10.1038/nrd2056.
4. Rusilowicz-Jones EV, Urbé S, Clague MJ. Protein degradation on the global scale. *Mol Cell* 2022;**82**:1414–23. https://doi.org/10.1016/j.molcel.2022.02.027.
5. Morreale FE, Kleine S, Leodolter J. *et al.* BacPROTACs mediate targeted protein degradation in bacteria. *Cell* 2022;**185**:2338–2353.e2318. https://doi.org/10.1016/j.cell.2022.05.009.
6. Lee JM, Hammarén HM, Savitski MM. *et al.* Control of protein stability by post-translational modifications. *Nat Commun* 2023;**14**:201. https://doi.org/10.1038/s41467-023-35795-8.
7. Hanna J, Guerra-Moreno A, Ang J. *et al.* Protein degradation and the pathologic basis of disease. *Am J Pathol* 2019;**189**:94–103. https://doi.org/10.1016/j.ajpath.2018.09.004.
8. Li X, Pu W, Zheng Q. *et al.* Proteolysis-targeting chimeras (PROTACs) in cancer therapy. *Mol Cancer* 2022;**21**:99. https://doi.org/10.1186/s12943-021-01434-3.
9. Ravid T, Hochstrasser M. Diversity of degradation signals in the ubiquitin–proteasome system. *Nat Rev Mol Cell Biol* 2008;**9**:679–89. https://doi.org/10.1038/nrm2468.
10. Bence NF, Sampat RM, Kopito RR. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* 2001;**292**:1552–5. https://doi.org/10.1126/science.292.5521.1552.
11. Xu H-D, Liang R-P, Wang Y-G. *et al.* mUSP: A high-accuracy map of the in situ crosstalk of ubiquitylation and SUMOylation proteome predicted via the feature enhancement approach. *Brief Bioinform* 2021;**22**:bbaa050. https://doi.org/10.1093/bib/bbaa050.
12. Varshavsky A. N-degron and C-degron pathways of protein degradation. *Proc Natl Acad Sci U S A* 2019;**116**:358–66. https://doi.org/10.1073/pnas.1816596116.
13. Koren I, Timms RT, Kula T. *et al.* The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* 2018;**173**:e1614. https://doi.org/10.1016/j.cell.2018.04.028.
14. Mészáros B, Kumar M, Gibson TJ. *et al.* Degrons in cancer. *Sci Signal* 2017;**10**:eaak9982. https://doi.org/10.1126/scisignal.aak9982.
15. Pan M, Zheng Q, Wang T. *et al.* Structural insights into Ubr1-mediated N-degron polyubiquitination. *Nature* 2021;**600**:334–8. https://doi.org/10.1038/s41586-021-04097-8.
16. Ji CH, Kim HY, Heo AJ. *et al.* The N-degron pathway mediates ER-phagy. *Mol Cell* 2019;**75**:e1059. https://doi.org/10.1016/j.molcel.2019.06.028.
17. Inobe T, Fishbain S, Prakash S. *et al.* Defining the geometry of the two-component proteasome degron. *Nat Chem Biol* 2011;**7**:161–7. https://doi.org/10.1038/nchembio.521.
18. Sherpa D, Chrustowicz J, Schulman BA. How the ends signal the end: Regulation by E3 ubiquitin ligases recognizing protein termini. *Mol Cell* 2022;**82**:1424–38. https://doi.org/10.1016/j.molcel.2022.02.004.
19. Lucas X, Ciulli A. Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Curr Opin Struct Biol* 2017;**44**:101–10. https://doi.org/10.1016/j.sbi.2016.12.015.
20. Ichikawa S, Flaxman HA, Xu W. *et al.* The E3 ligase adapter cereblon targets the C-terminal cyclic imide degron. *Nature* 2022;**610**:775–82. https://doi.org/10.1038/s41586-022-05333-5.
21. Inuzuka H, Tseng A, Gao D. *et al.* Phosphorylation by casein kinase I promotes the turnover of the Mdm2 oncoprotein via the SCFβ-TRCP ubiquitin ligase. *Cancer Cell* 2010;**18**:147–59. https://doi.org/10.1016/j.ccr.2010.06.015.
22. Nihira NT, Ogura K, Shimizu K. *et al.* Acetylation-dependent regulation of MDM2 E3 ligase activity dictates its oncogenic function. *Sci Signal* 2017;**10**:eaai8026. https://doi.org/10.1126/scisignal.aai8026.
23. Skaar JR, Pagan JK, Pagano M. SCF ubiquitin ligase-targeted therapies. *Nat Rev Drug Discov* 2014;**13**:889–903. https://doi.org/10.1038/nrd4432.
24. Yeh CW, Huang WC, Hsu PH. *et al.* The C-degron pathway eliminates mislocalized proteins and products of deubiquitinating enzymes. *EMBO J* 2021;**40**:e105846. https://doi.org/10.15252/embj.2020105846.
25. Makaros Y, Raiff A, Timms RT. *et al.* Ubiquitin-independent proteasomal degradation driven by C-degron pathways. *Mol Cell* 2023;**83**:e1927. https://doi.org/10.1016/j.molcel.2023.04.023.
26. Zhang Z, Sie B, Chang A. *et al.* Elucidation of E3 ubiquitin ligase specificity through proteome-wide internal degron mapping. *Mol Cell* 2023;**83**:e3376. https://doi.org/10.1016/j.molcel.2023.08.022.
27. Timms RT, Mena EL, Leng Y. *et al.* Defining E3 ligase–substrate relationships through multiplex CRISPR screening. *Nat Cell Biol* 2023;**25**:1535–1545. https://doi.org/10.1038/s41556-023-01229-2.
28. Tokheim C, Wang X, Timms RT. *et al.* Systematic characterization of mutations altering protein degradation in human cancers. *Mol Cell* 2021;**81**:1292–1308.e1211. https://doi.org/10.1016/j.molcel.2021.01.020.
29. Lin H-C, Yeh C-W, Chen Y-F. *et al.* C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol Cell* 2018;**70**:e603. https://doi.org/10.1016/j.molcel.2018.04.006.
30. Hou C, Li Y, Wang M. *et al.* Systematic prediction of degrons and E3 ubiquitin ligase binding via deep learning. *BMC Biol* 2022;**20**:162. https://doi.org/10.1186/s12915-022-01364-6.
31. Szulc NA, Stefaniak F, Piechota M. *et al.* DEGRONOPEDIA: A web server for proteome-wide inspection of degrons. *Nucleic Acids Res* 2024;**52**:W221–W232. https://doi.org/10.1093/nar/gkae238.
32. Martínez-Jiménez F, Muiños F, López-Arribillaga E. *et al.* Systematic analysis of alterations in the ubiquitin proteolysis system reveals its contribution to driver mutations in cancer. *Nat Cancer* 2020;**1**:122–35. https://doi.org/10.1038/s43018-019-0001-2.
33. He J, Chao WC, Zhang Z. *et al.* Insights into degron recognition by APC/C coactivators from the structure of an Acm1-Cdh1 complex. *Mol Cell* 2013;**50**:649–60. https://doi.org/10.1016/j.molcel.2013.04.024.
34. Kumar M, Michael S, Alvarado-Valverde J. *et al.* ELM-the eukaryotic linear motif resource-2024 update. *Nucleic Acids Res* 2024;**52**:D442–d455. https://doi.org/10.1093/nar/gkad1058.
35. Xu H, Hu R, Zhao Z. DegronMD: Leveraging evolutionary and structural features for deciphering protein-targeted degradation, mutations, and drug response to Degrons. *Mol Biol Evol* 2023;**40**:msad253. https://doi.org/10.1093/molbev/msad253.
36. Kumar M, Gouw M, Michael S. *et al.* ELM—The eukaryotic linear motif resource in 2020. *Nucleic Acids Res* 2020;**48**:D296–306. https://doi.org/10.1093/nar/gkz1030.
37. Guharoy M, Bhowmick P, Sallam M. *et al.* Tripartite degrons confer diversity and specificity on regulated protein degradation

in the ubiquitin-proteasome system. *Nat Commun* 2016;**7**:10239. https://doi.org/10.1038/ncomms10239.

38. Erdős G, Pajkos M, Dosztányi Z. IUPred3: Prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* 2021;**49**:W297–w303. https://doi.org/10.1093/nar/gkab408.
39. Yang Y, Heffernan R, Paliwal K. *et al.* SPIDER2: A package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. *Methods Mol Biol* 2017;**1484**:55–63. https://doi.org/10.1007/978-1-4939-6406-2_6.
40. Cilia E, Pancsa R, Tompa P. *et al.* The DynaMine webserver: Predicting protein dynamics from sequence. *Nucleic Acids Res* 2014;**42**:W264–70. https://doi.org/10.1093/nar/gku270.
41. Mistry J, Chuguransky S, Williams L. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* 2021;**49**:D412–d419. https://doi.org/10.1093/nar/gkaa913.
42. Xu H, Zhou J, Lin S. *et al.* PLMD: An updated data resource of protein lysine modifications. *J Genet Genomics* 2017;**44**:243–50. https://doi.org/10.1016/j.jgg.2017.03.007.
43. Lin S, Wang C, Zhou J. *et al.* EPSD: A well-annotated data resource of protein phosphorylation sites in eukaryotes. *Brief Bioinform* 2021;**22**:298–307. https://doi.org/10.1093/bib/bbz169.
44. Xu H, Zhao Z. NetBCE: An interpretable deep neural network for accurate prediction of linear B-cell epitopes. *Genomics Proteomics Bioinformatics* 2022;**20**:1002–12. https://doi.org/10.1016/j.gpb.2022.11.009.
45. Wang C, Xu H, Lin S. *et al.* GPS 5.0: An update on the prediction of kinase-specific phosphorylation sites in proteins. *Genomics Proteomics Bioinformatics* 2020;**18**:72–80. https://doi.org/10.1016/j.gpb.2020.01.001.
46. Mészáros B, Simon I, Dosztányi Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 2009;**5**:e1000376. https://doi.org/10.1371/journal.pcbi.1000376.
47. Lee TJ, Pouliot Y, Wagner V. *et al.* BioWarehouse: A bioinformatics database warehouse toolkit. *BMC Bioinformatics* 2006;**7**:170. https://doi.org/10.1186/1471-2105-7-170.
48. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Association for Computing Machinery, New York, NY, USA, 2016;785–794.
49. Heinzinger M, Elnaggar A, Wang Y. *et al.* Modeling aspects of the language of life through transfer-learning protein sequences. *BMC Bioinformatics* 2019;**20**:1–17. https://doi.org/10.1186/s12859-019-3220-8.
50. Xu H, Hu R, Jia P. *et al.* 6mA-finder: A novel online tool for predicting DNA N6-methyladenine sites in genomes. *Bioinformatics* 2020;**36**:3257–9. https://doi.org/10.1093/bioinformatics/btaa113.
51. Franz M, Lopes CT, Fong D. *et al.* Cytoscape.js 2023 update: A graph theory library for visualization and analysis. *Bioinformatics* 2023;**39**:btad031. https://doi.org/10.1093/bioinformatics/btad031.
52. Rego N, Koes D. 3Dmol.Js: Molecular visualization with WebGL. *Bioinformatics* 2015;**31**:1322–4. https://doi.org/10.1093/bioinformatics/btu829.
53. Jehl P, Manguy J, Shields DC. *et al.* ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Res* 2016;**44**:W11–5. https://doi.org/10.1093/nar/gkw265.
54. Becht E, McInnes L, Healy J. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2019;**37**:38–44. https://doi.org/10.1038/nbt.4314.
55. Schapira M, Calabrese MF, Bullock AN. *et al.* Targeted protein degradation: Expanding the toolbox. *Nat Rev Drug Discov* 2019;**18**:949–63. https://doi.org/10.1038/s41573-019-0047-y.
56. Chamberlain PP, Hamann LG. Development of targeted protein degradation therapeutics. *Nat Chem Biol* 2019;**15**:937–44. https://doi.org/10.1038/s41589-019-0362-y.
57. Zhao L, Zhao J, Zhong K. *et al.* Targeted protein degradation: Mechanisms, strategies and application. *Signal Transduct Target Ther* 2022;**7**:113. https://doi.org/10.1038/s41392-022-00966-4.
58. Dale B, Cheng M, Park K-S. *et al.* Advancing targeted protein degradation for cancer therapy. *Nat Rev Cancer* 2021;**21**:638–54. https://doi.org/10.1038/s41568-021-00365-x.
59. Samarasinghe KT, Crews CM. Targeted protein degradation: A promise for undruggable proteins. *Cell Chem Biol* 2021;**28**:934–51. https://doi.org/10.1016/j.chembiol.2021.04.011.
60. Békés M, Langley DR, Crews CM. PROTAC targeted protein degraders: The past is prologue. *Nat Rev Drug Discov* 2022;**21**:181–200. https://doi.org/10.1038/s41573-021-00371-6.
61. Dixon T, MacPherson D, Mostofian B. *et al.* Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry. *Nat Commun* 2022;**13**:5884. https://doi.org/10.1038/s41467-022-33575-4.