2024 Volume 2, Issue 2
Big Data Analytics for Healthcare
ISSN: 2630-4767 (Online)

Research on Bioinformatics and Molecular Simulation in Proteolysis Targeting Chimeras (PROTAC)

Jie Zhu[1], Jinyi Nie[1], Xiaoyi Liu[1], Zhongjie Liang[1,2]*

[1]Suzhou Medical College of Soochow University, Suzhou 215123, China

[2]Jiangsu Province Engineering Research Center of Precision Diagnostics and Therapeutics Development, Soochow University, Suzhou 215123, China

***Corresponding author:** Zhongjie Liang, zjliang@suda.edu.cn

Abstract:

Proteolysis targeting chimera (PROTAC) is a drug discovery strategy using a ubiquitin-proteasome system (UPS) to degrade the target protein. Unlike traditional small molecule drugs utilizing occupancy-driven pharmacology as the mode of action (MOA) to regulate protein activity, PROTACs function through forming stable target protein-PROTAC-E3 ubiquitin ligase ternary complex and use the ubiquitin-proteasome system to degrade the target protein. However, only a few E3 ubiquitin ligases have been used in PROTAC drug design now, and the space of target proteins that PROTAC can target needs to be further expanded. On the other hand, the complicated system of ternary crystal structures is difficult to capture and identify, computational simulation provides modeling of PROTAC-mediated ternary complex formation with effective approaches. Because of this, this review describes the recent progress of bioinformatics in expanding the landscape of E3 ubiquitin ligases and target proteins and summarizes the methods of computation simulation in modeling PROTAC ternary complex. Finally, the trend of development about PROTAC is prospected.

Keywords:

Proteolysis targeting chimera
E3 ubiquitin ligase
Target landscape
Ternary complex computational simulation
Protein translational modification
Heterobifunctional molecule

Online publication: December 27, 2024

1. The principles and characteristics of PROTAC

The concept of proteolysis targeting chimeras (PROTAC) was first proposed by the Crews and Deshaies laboratories in 2001. This team successfully achieved the degradation of the target protein methionine aminopeptidase-2 (MetAp-2) using PROTAC technology [1]. Since then, more PROTACs have entered clinical studies. Among them, PROTAC targeting the degradation of androgen receptor (AR) has entered Phase II clinical trials, and PROTAC targeting estrogen receptor (ER) is about to enter Phase III clinical trials, for the treatment of prostate

cancer and breast cancer, respectively [2]. So far, PROTAC technology has been used to degrade different target proteins related to various diseases, showing good clinical efficacy in cancer, immune diseases, neurodegenerative diseases, and cardiovascular diseases [3].

PROTAC is a bifunctional molecule that targets the protein of interest (POI) and recruits ubiquitin ligases (E3), achieving targeted degradation of the POI. PROTAC molecules consist of three parts: a ligand that binds to the target protein, a ligand that binds to the E3 ubiquitin ligase, and a linker that connects the two ligands. By inducing the proximity of E3 ligase and the target protein to form a ternary complex, PROTAC utilizes the ubiquitin-proteasome system (UPS) to degrade the target protein [4] (**Figure 1A**). In eukaryotic cells, the ubiquitin-proteasome system is the primary mechanism for maintaining protein homeostasis by degrading defective and damaged proteins. In this pathway, proteins are recognized by the proteasome and targeted for degradation through a three-step process involving three enzymes: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3). First, the activated ubiquitin (Ub) molecule is connected to the E1 ubiquitin-activating enzyme. Then, E1 binds to the E2 ubiquitin-conjugating enzyme and transfers the activated ubiquitin to E2. Finally, the E3 ubiquitin ligase catalyzes the transfer of ubiquitin from E2 to the lysine side chain of the target protein [5].

Compared to other drugs, PROTAC offers advantages such as targeting undruggable proteins, prolonged duration of action, reusability, overcoming drug resistance, high selectivity and specificity, and multiple administration routes [6] (**Figure 1B**), indicating a broad application prospect. Unlike the "occupancy-driven" mode of action of traditional drugs that require long-term and high-intensity binding to active sites, PROTAC degrades target proteins in an "event-driven" manner. PROTAC only needs to bind to induce target

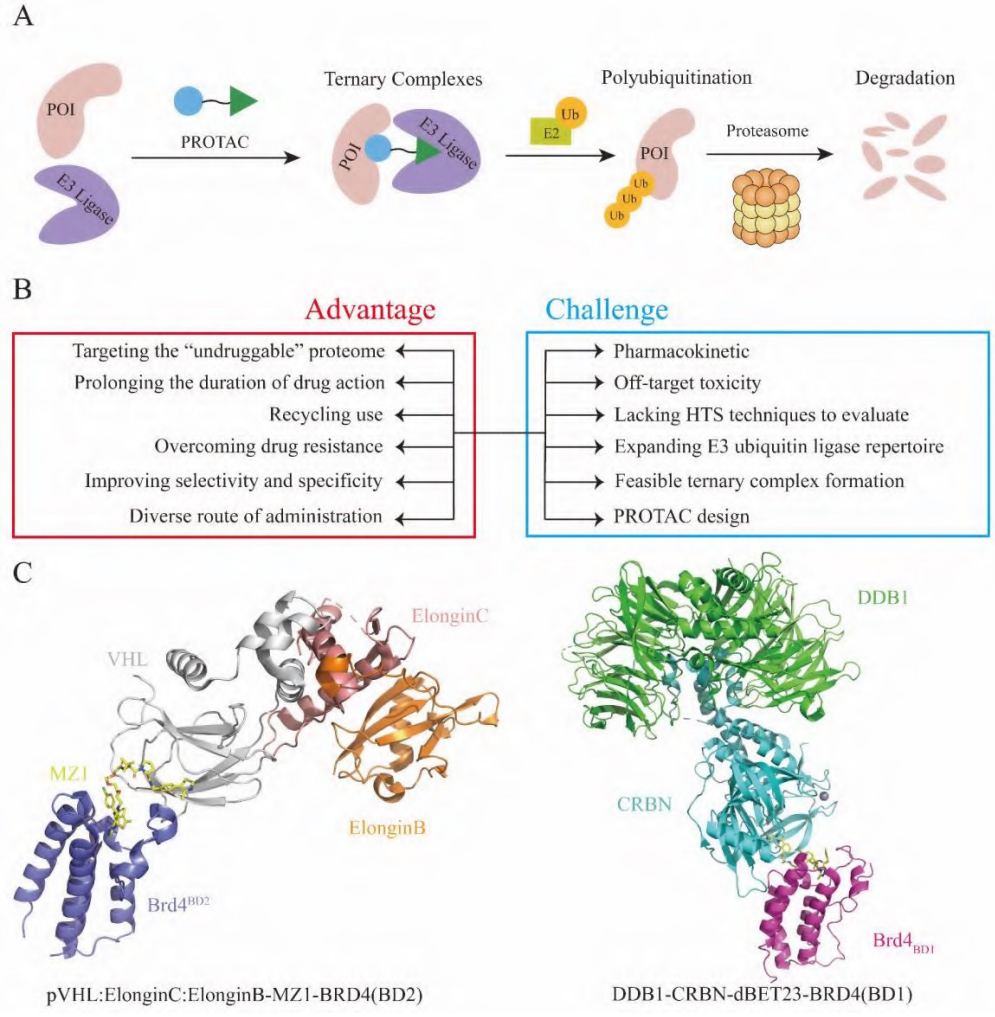

Figure 1. Proteolysis targeting chimeras (PROTAC). (A) The mechanism of PROTAC; (B) Advantage and challenge of PROTAC; (C) Crystal structures of $Brd4^{BD2}$:MZ1:VHL-ElonginC-ElonginB complex (PDB ID: 5T35, left) and DDB1ΔB-CRBN-dBET23-$BRD4_{BD1}$ complex (PDB ID: 6BN7, right). POI: Protein of interest.

protein degradation, greatly expanding the space of targetable proteins. Even if PROTAC dissociates, the ubiquitinated target protein continues to be degraded, and it takes a long time for the target protein to return to a functional level. PROTAC can be recycled and reused in the next degradation cycle with unchanged activity [7], avoiding high-dose administration and corresponding adverse reactions, allowing PROTAC to achieve higher activity at lower concentrations. Preclinical studies have shown preliminary data indicating the potential of this low-dose, low-frequency dosing approach. Furthermore, while traditional small molecule inhibitors can only block protein activity, PROTAC can eliminate the target protein, blocking not only its catalytic function but also affecting other functions such as protein-protein interactions and the ability to form large protein complexes. In a way, PROTAC can address acquired resistance caused by traditional inhibitors because target protein degradation may block certain feedback mechanisms of cellular protein homeostasis [6]. Although studies have shown that PROTAC treatment may also lead to drug resistance, this resistance mechanism is different from that of small molecule inhibitors and is due to mutations in E3 ligase that limit the formation of ternary complexes. This type of resistance can be overcome by modifying the E3 ligase.

Compared to macromolecular drugs such as monoclonal antibodies, PROTACs can target a large number of intracellular proteins and exhibit high tissue permeability, enabling oral administration. Depending on the disease and requirements, PROTAC protein degraders can be developed for oral, injection, and infusion administration. Indeed, PROTAC technology has achieved significant success over the past few decades, providing new avenues for drug development. However, this new technology faces several challenges, including issues related to PROTAC pharmacokinetics, limited target protein localization, improving selectivity and specificity, and off-target effects [6]. To address these challenges, increasing research efforts have focused on developing new derivative technologies. For instance, antibody-PROTAC conjugates (APCs) have been developed to conjugate PROTACs to tumor-specific antibodies [8], enabling selective targeting of proteins located in specific tissue cells in a tumor-directed manner to achieve specific targeted degradation. Additionally, trivalent PROTACs have been designed to enhance molecular affinity and synergism, forming stable complexes that better bind to target proteins, improve protein degradation efficiency, and exhibit more potent anticancer activity, particularly against tumors that have developed drug resistance [9]. Furthermore, to overcome the limitation of PROTACs in degrading proteins at restricted locations, researchers have devised novel techniques that utilize lysosomal degradation pathways to degrade membrane and transmembrane proteins, enabling targeted localization of different target proteins [10].

Despite the development of many new technologies to compensate for PROTAC's inherent limitations, this technology still faces several non-negligible challenges (**Figure 1B**). Firstly, the ADMET properties of PROTACs need optimization. PROTACs themselves do not conform to the rule of five for drug-likeness, with molecular weights ranging from 700 to 1,200 Da, and poor solubility, membrane permeability, and oral bioavailability [6]. Secondly, there is a possibility of off-target effects during PROTAC-induced target protein degradation, which can potentially cause greater toxicity than traditional drugs, and there is a lack of effective detection methods. Although derivative technologies such as photochemical targeting chimeras (PHOTACs), semiconducting polymer nano-PROTACs (SPNpros), and folate-group-linked PROTACs (floate-PROTACs) have been developed to address the issue of off-target toxicity [11], they are not applicable in all cases. Furthermore, there is currently no effective high-throughput screening technology for rapidly and efficiently evaluating PROTAC's ability to degrade target proteins, which slows down the speed and success rate of PROTAC development.

Additionally, E3 ligases are one of the key factors determining the selectivity and specificity of target protein degradation, which is crucial for achieving precise targeted protein degradation. Currently, two E3 ligases, CRBN (Cereblon) and VHL (Von-Hippel-Lindau), are primarily selected for clinical development. However, mutations in E3 ligases can affect degradation efficacy, highlighting the need to expand the space of E3 ligases to improve PROTAC's selectivity and specificity in targeting proteins [12]. In the process of PROTAC-induced target protein degradation, the formation of a ternary complex consisting of the target protein, PROTAC, and E3 ubiquitin ligase is

critical for achieving effective and selective degradation of the target protein. For PROTACs, although the target protein-binding ligand and E3 ligase ligand do not need to bind tightly to proteins like other drugs, both ligands must have a certain affinity for the target protein and E3 ligase. When the concentration of PROTAC increases to a certain level, excess PROTAC can form binary complexes with either the target protein or E3 ligase, a phenomenon known as the hook effect [12]. This is unfavorable for the formation of stable ternary complexes and may lead to severe adverse reactions and increased off-target toxicity. Furthermore, the interaction between the target protein and E3 ligase can also affect the formation of the ternary complex. Attractive interactions can stabilize the ternary complex, while repulsive interactions can destabilize it, thereby influencing PROTAC's degradation efficiency. This is referred to as cooperativity, which is defined as the ratio of dissociation constants for PROTAC-bound binary and ternary complexes (α) and describes the efficiency of ternary complex formation [13]. The first crystal structure of a ternary complex (BRD4-MZ1-VHL, PDB ID: 5T35) revealed that PROTAC-induced surface electrostatic interactions between the target protein and E3 ligase play a positive cooperative role in stabilizing the ternary complex [14]. A higher α is believed to be associated with a weaker hook effect, which is a crucial aspect of PROTAC design and development. Although studies have shown that ternary complexes with negative cooperativity can also lead to degradation, cooperativity is generally important for assessing PROTAC-induced ternary complex formation. Understanding the structure and stability of ternary complexes aids in optimizing degradation and improving PROTAC design [15]. However, only a few crystal structures of ternary complexes have been resolved, such as the ternary complex crystal structures of different bromodomains of the BRD4 protein interacting with VHL and CRBN systems (**Figure 1C**) [14,16]. During complex formation, various factors such as PROTAC linker length, composition, and attachment position, ligand-protein interactions, and the mode of interaction between the two proteins can result in different binding modes. As current experimental methods are challenging to obtain crystal structures of ternary complexes, rational PROTAC modification is difficult without these structures.

In summary, expanding the space of E3 ligases and targets, as well as constructing ternary complexes, are significant for PROTAC development and design. Therefore, this article first introduces bioinformatics methods for expanding the space of E3 ligases and degradable target proteins, particularly focusing on PROTAC selectivity and specificity. Secondly, it presents computational simulation methods, including protein docking, sampling, clustering, and currently established computational pipelines for ternary complex modeling. Finally, it explores other new technologies based on the proximity effect, like PROTAC, and discusses the future development of bioinformatics and molecular simulation techniques to further advance the field of drug design.

2. Research on bioinformatics in expanding E3 ubiquitin ligase and target space

2.1. Research on bioinformatics in expanding E3 ubiquitin ligases

Many signaling pathways in cells selectively degrade certain key regulatory proteins through the ubiquitin-proteasome pathway. This highly specific recognition mechanism for specific proteins is primarily determined by E3 ubiquitin ligases. E3 ubiquitin ligases can catalyze degradation through various mechanisms, and more than 600 E3 ligases in the human genome are mainly classified into three categories based on different catalytic mechanisms (**Figure 2A**) [17]. The largest class of E3 ubiquitin ligases is the RING (really interesting new gene) ligase, which transfers ubiquitin to the lysine residue of the target protein by binding to the target protein and E2-Ub [17]. The second category is the HECT (homologous to E6-AP C-terminus) ligase, which transfers ubiquitin to the HECT domain by binding to E2-Ub. The catalytic domain's cysteine residue covalently binds to ubiquitin, forming a covalent thioester intermediate, and finally transfers ubiquitin to the lysine residue of the target protein [18]. The third category is the RBR (RING-between-RING) ligase, which combines the characteristics of RING and HECT ligases. It contains three domains: the RING1 domain that binds to E2-Ub, the RING2 domain that catalyzes ubiquitin transfer, and the IBR domain located between the two RING domains. RING1 recognizes E2-Ub first, then ubiquitin is

transferred to the catalytic cysteine residue of RING2 to form a thioester intermediate and finally transferred to the lysine residue of the target protein [19]. Different E3 ligase families are presented in the form of a phylogenetic tree based on the E3 ligases collected in the Ubihub database (**Figure 2B**) [20]. Unlike the three classifications mentioned above, the left side shows monomeric E3 ligase families, including HECT, RBR, TRIM (tripartite motif), TRAF (tumor necrosis factor receptor-associated factor), and other families. The right side represents complex E3 ligases that rely on multiple subunits, including families such as DCAF (DDB1-CUL4-associated factor), SOCS (suppressor of cytokine signaling), BTB (broad-complex, tramtrack, and bric-a-brac), FBXL (F-box with leucine-rich amino acid repeats), FBXW (F-box with WD40 amino acid repeats), FBXO (F-box only with uncharacterized domains), etc.

Although there are more than 600 E3 ligases, only a few have developed corresponding E3 ligase ligands [21]. E3 ligase ligands are one of the important determinants of PROTAC degradation efficiency, target protein selectivity, and specificity. Currently, the E3 ligases used for PROTAC molecule design are limited to VHL, CRBN, MDM2 (murine double minute 2), and IAPs (inhibitor of apoptosis) proteins. All four E3 ligases, VHL, CRBN, MDM2, and IAPs, belong to the RING family. Most PROTAC molecules entering clinical trials are developed based on CRBN ligands, and a few are based on VHL ligands. The heavy reliance on these two E3 ligases poses challenges to the clinical development of PROTACs. Mutations can rapidly lead to PROTAC resistance mechanisms and may even cause off-target effects resulting in severe adverse reactions [22]. Therefore, expanding the available E3 ubiquitin ligases for

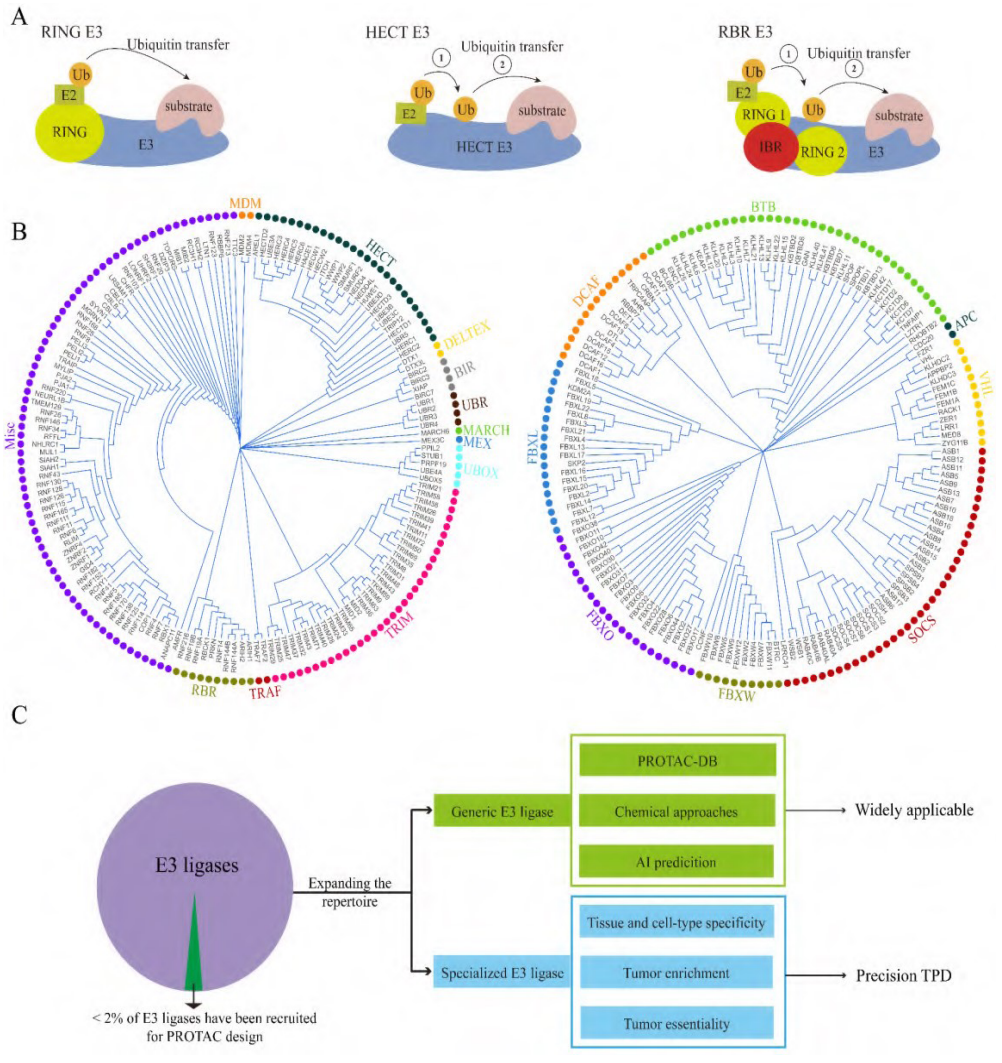

Figure 2. Expanding the arsenal of E3 ubiquitin ligases. (A) Classes of E3 ubiquitin ligases; (B) Phylogenetic trees of E3 ubiquitin ligases; (C) Expanding the ligase landscape. TPD: Target protein degradation.

PROTACs is a necessary approach to optimize targeted protein degradation strategies.

Expanding the available E3 ligases for PROTACs (**Figure 2C**) involves several steps. Firstly, researchers can search for broadly applicable E3 ligases, similar to CRBN and VHL, that can interact with multiple target proteins for different disease treatments. The PROTAC-DB database developed by Weng et al. [23] collects E3 ligases currently used in PROTAC molecule design. Secondly, besides the collected E3 ligases, researchers use chemical methods like activity-based protein profiling (ABPP) [24] to expand the library of E3 ligases for PROTAC development. PROTACs based on E3 ligase ligands such as ring finger protein 4 (RNF4), ring finger protein 114 (RNF114), DCAF16 (DDB1-Cul4-associated factor 16), and AhR (arylhydrocarbon receptor) have been developed [25]. Furthermore, in addition to the E3 ligases already proven for targeted protein degradation (TPD), some E3 ligases do not have accessible or complete crystal structures but can theoretically be used for PROTAC development. Modeling can predict possible structural models for these E3 ligases. In recent years, artificial intelligence (AI)-based predictions, such as those from Google/DeepMind [26] and RoseTTAFold [27], have provided excellent quality tertiary structure models. For instance, the AlphaFold database launched by DeepMind offers high-quality structural prediction models, opening possibilities for drug discovery for many target proteins and ligases with unknown structures until now [21].

Although discovering universally applicable E3 ligases for PROTAC molecular design represents a practical and valuable approach, enhancing their selectivity and specificity, and avoiding adverse reactions remain primary goals in drug development. Differences in degradation among various E3 ligases are determined by several factors, including the degree of shape complementarity between the ligase and target protein, the ability to form stable ternary complexes, subcellular localization disparities, and cell type-specific expression profiles of the ligase and target protein [21]. Among these, the characteristic expression profiles of E3 ligases play a crucial role in discovering new E3 ligases. Identifying E3 ligases with unique expression profiles can improve PROTAC selectivity, reduce cytotoxicity, and achieve precisely targeted protein degradation (**Figure 2C**). Currently, publicly available databases such as GTEx (genotype-tissue expression), TCGA (The Cancer Genome Atlas), and HPA (Human Protein Atlas) [28] can be utilized to analyze E3 ligase expression and identify those with distinct expression patterns [12]. So far, several E3 ligases with tissue-specific expression have been identified, including KLHL40 (Kelch-like family member 40) and KLHL41 (Kelch-like family member 41) enriched in skeletal muscle [29], and RNF182 (ring finger protein 182) enriched in the central nervous system [30]. Additionally, some E3 ligases demonstrate reverse specificity with low expression in certain tissues or cell types. For instance, Bcl-XL (B-cell lymphoma extra large) is an anti-apoptotic protein, but drugs targeting this protein may lead to reduced platelet levels. VHL, which has low expression in platelets, can be recruited using PROTAC to target specific tumor types without causing thrombocytopenia as an adverse reaction [31].

Beyond analyzing the tissue specificity of E3 ligases, investigating their tumor specificity or tumor enrichment aids in achieving precise targeted protein degradation (**Figure 2C**). Typically, the enrichment of E3 ligases aligns with the tumor's dependency on their expression. The CERES algorithm [33] (computational correction of copy-number effect in CRISPR-Cas9 essentiality screens) developed by DepMap [32] can analyze E3 ligase dependency scores across multiple tumor cell lines, measuring their importance for the growth and survival of specific tumor cells [34]. Tumor cells have a lower ability to develop resistance to PROTACs based on critical E3 ligases, but the toxicity of PROTACs developed from such E3 ligases remains to be tested. Furthermore, cancer-testis antigens (CTAs), a subset of E3 ligases, are restricted in normal testes but highly overexpressed in various cancers. The melanoma antigen (MAGE) family [35], often referred to as MAGE-RING ligases (MRLs), participate in recruiting substrates for RING E3 ligases and function as multi-subunit complexes [36]. Although not all MAGE E3 ligases are tumor-specific, such tissue and disease-specific E3 ligases represent an area for expanding E3 ligase space research [21].

PROTACs exert their effects by forming stable ternary complexes, enabling even ligands with low binding affinity to function effectively. Since the hook effect depends on the expression levels of E3 ligases and target proteins, PROTAC degradation efficiency can vary across different cell types and tissues [28]. Therefore,

specific E3 ligases are crucial for forming stable ternary complexes, reducing drug toxicity, and targeting more challenging proteins.

2.2. Research on bioinformatics in expanding the space of degradable targets

PROTAC technology has been successfully applied to degrade various target proteins, including kinases, nuclear receptors, epigenetic proteins, and misfolded proteins. Increasingly, target proteins are being confirmed as "PROTAC table," and can be used to design PROTACs for targeted degradation. PROTAC target proteins that have entered clinical trials include the androgen receptor (AR), estrogen receptor (ER), interleukin-1 receptor-associated kinase 4 (IRAK4), and signal transducer and activator of transcription 3 (STAT3) [6]. The PROTAC-DB database covers information on 3,270 PROTAC molecules, 365 target protein ligands, 82 E3 ligase ligands, 1,501 linkers, and 664 ternary complex models. PROTACs do not need to bind to the biologically active sites of target proteins, expanding the space of target proteins and enabling the targeted degradation of targets considered difficult to drug.

Statistics show that over 80% of human proteins are "undruggable" target proteins. Understanding whether these proteins can be targeted for degradation by PROTACs (PROTAC ability) is key. Schneider et al. (2021) [37] conducted the first systematic and comprehensive evaluation of potential drug targets from an omics perspective. They analyzed proteins that could serve as drug targets by integrating information from various publicly available data sources, collecting information on six aspects: clinical development, literature reports, ubiquitination, protein half-life (turnover), small-molecule binders, and location (**Figure 3**).

Based on six aspects of information, target proteins were systematically analyzed and classified into different categories.

(1) The ClinicalTrials.gov website was searched to determine whether the protein is a PROTAC drug target that has entered clinical trials, and it was categorized into three groups:
   (a) Target proteins in Phase IV clinical PROTAC trials;
   (b) Target proteins in Phase II or III clinical PROTAC trials;
   (c) Target proteins in Phase I clinical PROTAC trials.

(2) The PubMed database was searched to check if there were any reported PROTAC studies on the target protein.

(3) Target proteins were classified into two categories based on the collection of ubiquitination site information on the protein:
   (a) Proteins with ubiquitination site information provided by the UniProt database;
   (b) Proteins with ubiquitination sites provided by the PhosphoSitePlus database, mUbiSiDa database, and proteomics-based experimental methods [38].

(4) The half-life data of target proteins was collected, as proteins with very short half-lives may limit the degradation efficiency of PROTACs. Mathieson *et al.* (2018) [39] used proteomics based on stable isotope labeling with amino acids in cell culture (SILAC) to determine the range of protein half-lives in various cell types.

(5) It was investigated whether the target protein had small molecule ligands. Small molecule ligands for target proteins only require appropriate binding affinity, and even transient interactions are sufficient to exert PROTAC degradation

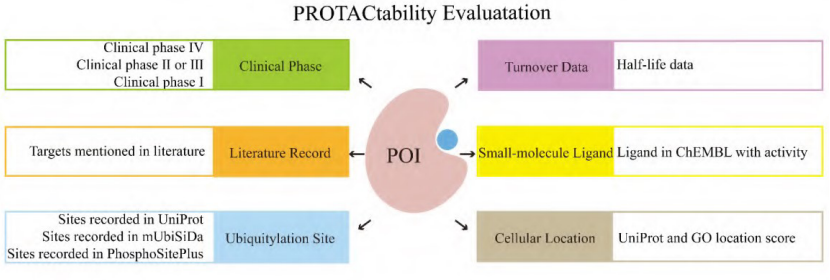

Figure 3. Expanding the PROTAC table target proteins.

function. In target-based activity assays reported in the ChEMBL database, the measured activity of small molecule ligands is at least 10 $\mu molL^{-1}$.

(6) Protein location data was obtained by searching the UniProt database and GO database, and a location score was assigned to each target protein. Proteins located in the cytoplasm or nucleus are considered to be in favorable locations, membrane proteins are considered to be in gray locations, and the rest are considered to be in unfavorable locations. These locations were then further classified based on confidence level to obtain a location score.

Based on the above six types of information for target proteins, targets can be further simplified into four PROTAC ability categories: clinical priority (presence of clinical data), literature priority (reported in the literature), degradable opportunity (meets at least one of the three criteria: has ubiquitination sites, has half-life data, and has reported small molecule ligands, and has a good location score), and incomplete evidence (insufficient evidence for degradation). The first three categories of target proteins are considered "PROTAC table" and can be evaluated and quantified for their potential as degradable targets [37].

3. Research on computational simulation methods in PROTAC ternary complex structure modeling

3.1. Computational analysis methods for constructing PROTAC ternary complex structure models

The formation of the "target protein-PROTAC-E3 ligase" ternary complex is crucial for PROTAC-induced target protein degradation. Effective and specific degradation of the target protein can only occur when PROTAC simultaneously binds to both the target protein and E3 ligase, forming a stable ternary complex. However, the complex crystal structure of the ternary complex is often difficult to characterize. In recent decades, biophysical methods have provided powerful tools to describe the formation of ternary complexes. To better understand the structural basis of PROTAC ternary complex formation, molecular simulation can be used to study PROTAC selectivity for target proteins and provide molecular-level explanations for degradation activity [12]. Currently, computational methods used to construct PROTAC ternary complexes mainly include structure generation and analysis (**Figure 4**) [13]. Structure generation primarily

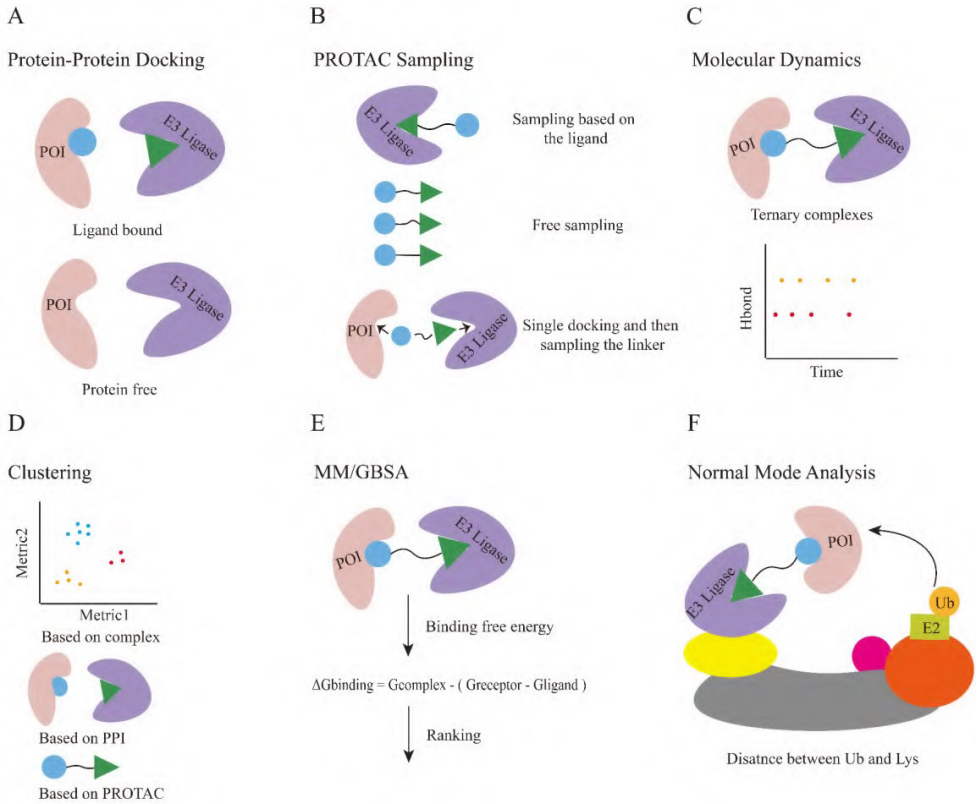

Figure 4. Computational methods applied in target protein degradation (TPD). (A) Protein-protein docking; (B) PROTAC conformation sampling; (C) Molecular dynamics simulation; (D) Clustering; (E) MM/GBSA; (F) Normal mode analysis.

Table 1. Characteristics of the key computational methods applied in TPD

Model generation technique		Analysis technique	
Protein-protein docking	Accurate prediction of binary protein-protein complexes	Clustering	Categorize the conformational space to get representative poses
PROTAC conformational sampling	Sample PROTAC conformations restrained to the protein-protein complex	Estimating energies with MM/GBSA	Improve the ranking accuracy of ternary complexes and estimate the stability and cooperativity
Molecular dynamics (MD)	Obtain the conformational ensemble of ternary complex	Normal mode analysis	Describe the collective motion of the protein complexes

involves protein-protein docking, PROTAC sampling, and molecular dynamics simulation. The analysis mainly includes clustering, MM/GBSA (molecular mechanics/generalized born surface area), and normal mode analysis (**Table 1**) [13,40].

3.1.1. Protein-protein docking methods

To obtain a reasonable ternary complex model, the structural information of proteins must first be acquired, determining whether crystal structures are available and whether they bind to ligands. After identifying the protein structures to be used, the target protein and E3 ligase are docked to obtain a structural model of protein interaction. Then, based on this, a PROTAC-bound ternary complex is generated (**Figure 4A**). However, not all of these protein interaction structural models are reasonable, and they need to be scored and evaluated. The best structure is then selected for optimization. Schiedel *et al.* (2018) [41] were the first to use the protein docking software HADDOCK [42] in the study of PROTAC-induced target protein degradation. Later, more protein-protein docking tools were developed to obtain reasonable protein docking conformations. In recent years, Rosetta Dock [43], based on the Monte Carlo algorithm, has become one of the most widely used protein-protein docking tools in TPD. Nowak *et al.* (2018) [16] used Rosetta Dock to dock CRBN and the binding domain 1 (BD1) of the bromodomain BRD4, generating 20,000 docking structures. Among the 200 lowest-energy conformations, they identified one that was very similar to the complex crystal structure. They designed a PROTAC by calculating the shortest distance between the solvent-exposed atom pairs of the BRD4 ligand JQ1 and the CRBN ligand lenalidomide in the top 200 conformations. Other commonly used protein docking tools include ZDOCK [44], HDOCK [45], ClusPro [46], PatchDock [47], and LightDock [48].

3.1.2. PROTAC conformation sampling methods

Protein-protein docking yields possible binding conformations between the target protein and E3 ligase in the ternary complex, while in reality, the ternary complex is formed through PROTAC induction. To determine the exact conformation of PROTAC binding to the two proteins, PROTAC sampling needs to be performed within the corresponding protein docking conformations, yielding a protein-protein structure suitable for PROTAC binding. PROTAC conformation sampling methods include:

(1) Sampling PROTAC based on the ligand naturally bound to the protein;
(2) Directly sampling the entire PROTAC molecule;
(3) Separately docking the protein and ligand, and generating the complete PROTAC molecule through the linker (**Figure 4B**).

Both Zaidman *et al.* (2020) [49] and Bai *et al.* (2021) [50] generated linker conformations based on protein-protein interactions to obtain PROTAC conformation sets. Currently, individual sampling is mainly used to acquire PROTAC conformations, including separate sampling of the linker and sampling of the entire PROTAC based on chemical information. PROTAC docking can generate and score geometric shapes around protein binding sites, or use docking scoring functions to evaluate the binding energy of PROTAC conformations. The combination of protein-protein docking and PROTAC conformation sampling methods can predict structural models of potential ternary complexes [51]. The commonly used conformation

generation software is RDKit, an open-source tool based on Python. Other commonly used software include OMEGA [52] and MOE.

3.1.3. Molecular dynamics simulation methods

Proteins are dynamic, and this characteristic is particularly important in predicting PROTAC ternary complexes. Molecular docking only obtains reasonable binding modes from the scoring function and conformational space sampling to determine whether binding can occur, but it cannot judge whether PROTAC can stably bind to the target protein and E3 ligase. Additionally, molecular docking cannot capture protein-ligand-induced fit effects or conformational changes between apo (monomer) and holo (complex) states [53]. Molecular dynamics (MD) simulation predicts how each atom in a protein or other molecular system moves with time based on general models of interatomic interactions in physics. MD reveals the positions of all atoms in a molecule with femtosecond time resolution, providing atomic-level explanations for important molecular processes [54]. Importantly, this simulation can also predict the response mechanisms of biomolecules at the atomic level to perturbations such as mutations, post-translational modifications, protonation, or ligand binding [55]. By performing MD simulations to observe the movement of PROTAC ternary complexes, the stability of the complexes can be evaluated [13]. Weerakoon *et al.* (2022) [56] studied the conformational changes of MZ1 and dBET6 using MD simulations and found that the conformational ensembles generated by MD results were consistent with those measured by nuclear magnetic resonance techniques. Commonly used molecular dynamics simulation programs include LAMMPS, AMBER, CHARMM, Tinker, NAMD, GROMACS, and OpenMM.

3.1.4. Clustering methods

Protein-protein docking, PROTAC conformational sampling, and molecular dynamics simulations generate a broad conformational space of target protein-PROTAC-E3 ligase ternary complexes, which requires further analysis. In this context, clustering can group the conformational ensemble of PROTAC ternary complexes into distinct clusters. PROTAC ternary complex conformations within the same cluster are similar, while conformations from different clusters exhibit significant differences. Representative conformations from different clusters can be selected for further analysis. The variation in clustering groups depends on the algorithm used. Commonly used clustering algorithms include K-means clustering, expectation-maximization clustering based on Gaussian mixture models, density-based spatial clustering of applications with noise, agglomerative clustering, and hierarchical clustering. Different clustering methods can be chosen based on the objective. Some computational modeling pipelines integrate clustering analysis after protein-protein docking methods, while other modeling tools perform clustering on ternary complex models to obtain more reasonable complex models. Weerakoon *et al.* (2022) [56] performed clustering based on MD simulation results and conducted a network analysis to visualize transitions between different states of the ternary complex based on the clustering. Therefore, clustering can categorize the conformational space to obtain stable representative conformations from simulations.

3.1.5. MM/GBSA method

Although clustering methods can provide grouping information for protein conformational spaces, other metrics like energy are more suitable for evaluating the stability of ternary complexes. Among various methods for assessing ligand-receptor interactions, the calculation of binding free energy plays a crucial role. Methods for calculating free energy include thermodynamic integration (TI), free energy perturbation (FEP), molecular mechanics/Poisson Boltzmann (Generalized Born) Surface Area (MM/PB(GB)SA), and linear interaction energy (LIE). Among them, MM/GBSA calculation is a commonly used method to estimate the binding free energy of small molecule ligands to biomacromolecules. Typically, based on MD simulation results, the binding free energy is decomposed into molecular mechanics terms and solvation energy for separate calculations. Liao *et al.* (2022) [57] developed an MD-based computational workflow that combines protein-protein docking with ligand docking to generate initial ternary complex confirmation candidates. MD simulations were performed on these structures to obtain stable ternary complex conformations, which were then clustered and ranked using MM/GBSA for energy scoring. Li *et al.* [40] proposed

a new strategy using MD to reorder ternary complexes generated by PRosettaC. The stability and synergy of PROTAC-induced ternary complexes were evaluated through MM/GBSA calculations. As protein-protein docking, PROTAC conformational sampling, and MD simulations produce a large number of conformations, MM/GBSA measures the degree of binding of ternary complex components by calculating energy and assessing the reasonability of ternary complex models.

3.1.6. Normal mode analysis method

Normal mode analysis (NMA) is widely used to predict vibrational modes in proteins, which are often associated with biological functions. Low-frequency motion modes correspond to global movements of the protein, while high-frequency motion modes correspond to local conformational changes. Studies have shown that these low-frequency motion modes correspond to functionally relevant global movements in proteins, and conformational transitions follow one or more of these normal modes [58]. NMA is often combined with coarse-grained models to simplify calculations [59], such as by using the anisotropic network model (ANM) to represent the protein structure as an elastic network of nodes and edges. Each node in the network represents an amino acid residue, and the edges represent interactions between them. Through this simplified elastic network analysis, normal modes can be used to study protein movements and obtain information about the global movement directions of residues [60]. By extracting information from the lowest frequency normal modes, ANM identifies flexible regions in the protein and their movement directions, predicting conformational changes in the protein.

By tracking changes in normal mode amplitudes with frequency, ANM can quantify the dynamic behavior of proteins and track large conformational changes. In studying the role of PROTACs, the ANM method analyzes the correlation between experimental degradation and data by calculating the distance between accessible lysine residues on E2Ub and the target protein [13].

In the simulation calculations for constructing target protein-PROTAC-E3 ligase ternary complexes, these modeling and analysis methods are often used interchangeably. Drummond *et al.* (2019) [61] proposed four methods for constructing target protein-PROTAC-E3 ligase ternary complexes, including:

(1) Generating a linker as a PROTAC-derived ternary complex based on two binding ligand proteins as the starting conformation, and then sampling the entire ternary complex;

(2) Independently sampling PROTAC conformations and then adding them to rigid proteins;

(3) Sampling PROTACs based on one of the proteins and then adding the second protein;

(4) Sampling PROTAC conformations but adding possible E3 ligase-target protein interaction structures through protein-protein docking.

Research experience has shown [16,49,50] that discovering favorable interactions between target proteins and E3 ligases is more meaningful for studying ternary complexes [15]. Sampling protein-protein and PROTAC or linker conformations independently is currently the main method for constructing ternary complex models.

In ternary complex studies, Schiedel *et al.* (2018) [41] obtained the binding conformation of the target protein Sirt2 and the E3 ligase CRBN through HADDOCK [42] docking, and then further evaluated the complex structure model using scoring functions, Van der Waals forces, and electrostatic interactions. The obtained conformations were clustered, and the three clusters with the highest scores were selected for further analysis. The molecular docking program GOLD was used to dock PROTAC into the Sirt2-CRBN complex to obtain docking conformations. MM/GBSA was used to analyze the obtained conformations, and low-energy conformations were selected for hydrogen bond analysis. The docking results showed that the binding mode of PROTAC to Sirt2-CRBN was consistent with the resolved crystal structures of Sirt2 and CRBN binding to small molecules, resulting in a reasonable Sirt2-PROTAC-CRBN ternary complex model. Crews *et al.* (2018) [62] used molecular docking to obtain a p38α-PROTAC-VHL ternary complex model, performed 120ns MD simulations, and hierarchical clustering to obtain representative conformations. They then further analyzed the conformation of PROTAC and the protein-protein interactions between VHL and p38α to investigate which interactions are favorable for forming stable ternary complexes. In another study from the Crews laboratory

[63], the degradation of p38δ, another subtype of the p38 MAPK family, was investigated based on the study of targeted degradation of p38α. Similarly, the corresponding ternary complex model was obtained through molecular docking, and short MD simulations were performed. By analyzing the structural differences between the two ternary complexes, the selective degradation of PROTAC molecules targeting different p38 subtypes was explained, guiding the design and optimization of PROTAC molecules. Experimental methods for studying target protein-PROTAC-E3 ligase ternary complexes mainly include biophysical methods for characterizing ternary complex formation and methods for determining ternary complex structures. Biophysical methods mainly include time-resolved fluorescence energy transfer (TR-FRET), amplified luminescent proximity homogeneous assay linked immunosorbent assay (AlphaLISA), fluorescence polarization (FP), isothermal titration calorimetry (ITC), and surface plasmon resonance (SPR). *In vitro* biophysical methods such as TR-FRET, AlphaLISA, FP, ITC, and SPR can characterize affinity, thermodynamic, and kinetic information during ternary complex formation. In live cell detection, nanobioluminescent resonance energy transfer (NanoBRET) and NanoLuc binary technology (NanoBiT) can dynamically monitor the formation of ternary complexes in live cells. Computational simulations involve performing molecular dynamics simulations on ternary complexes and combining them with MM/GBSA for binding energy calculations, which can predict PROTAC binding energies and the synergies of ternary complex formation, improving the prediction accuracy of ternary complex structure models. Methods for determining ternary complex structures mainly include experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM). Computational method workflows include protein-protein docking, PROTAC conformation sampling, molecular dynamics simulations, and other steps to obtain the conformational space of ternary complex structures. By structurally modeling known ternary complex crystal structures, the rationality of computational methods for ternary complex structure modeling can be evaluated [13].

3.2. Computational workflow for the structure modeling of PROTAC ternary complexes

Currently, there are mainly four computational methods used for PROTAC ternary complex research:

(1) Based on docking to obtain ternary complex models, such as the four methods proposed by Drummond *et al.* (2019) [61], PRosettaC proposed by Zaidman *et al.* (2020) [49], and PROTAC-Model proposed by Hou Tingjun *et al.* (2021) [51].

(2) Combines MD simulations to study ternary complexes, as demonstrated by Li *et al.* (2022) [40] who presented a strategy to reorder ternary complexes generated by PRosettaC based on MD.

(3) Studying ternary complexes based on ubiquitination models. For instance, Bai *et al.* (2022) [64] simulated the conformation of the CRL4A ligase complex and classified the ternary complexes into productive and unproductive complexes based on the distance between ubiquitin and lysine on the target protein.

(4) Utilizes machine learning to predict ternary complexes, as exemplified by DeepPROTACs [65], which uses machine learning to predict the targeted degradation ability of PROTAC small molecules.

Among these methods, PRosettaC [49] and PROTAC-Model [51] integrate RosettaDock [43] to provide online predictions of ternary complex models.

3.2.1. PRosettaC

Zaidman *et al.* (2020) [49] developed PRosettaC (**Figure 5A**), a tool based on the molecular modeling software Rosetta [43] to construct PROTAC-mediated ternary complexes. Firstly, PROTAC conformations are randomly generated using RDKit, and a distance threshold between the two ends of the protein-binding ligand is provided to obtain a distance distribution. Then, within the limits of this distance threshold, the PatchDock [47] software is used to filter the conformational space obtained from protein docking, significantly reducing the number of docked conformations. Next, RosettaDock [43] software is employed for local docking to generate 50 high-resolution models. Based on these protein docking models, the complete PROTAC is constructed. The PROTAC ternary complex models are obtained and analyzed using Rosetta to filter out high-energy complexes, selecting

the top 200 models with the highest scores. Finally, these ternary complexes are clustered, and the representative conformations of the clusters are considered the complex models closest to the crystal structure. This method restricts the search space for protein-protein docking by using the distance distribution of PROTAC and then limits the search space for ligand conformations through energy optimization and clustering techniques.

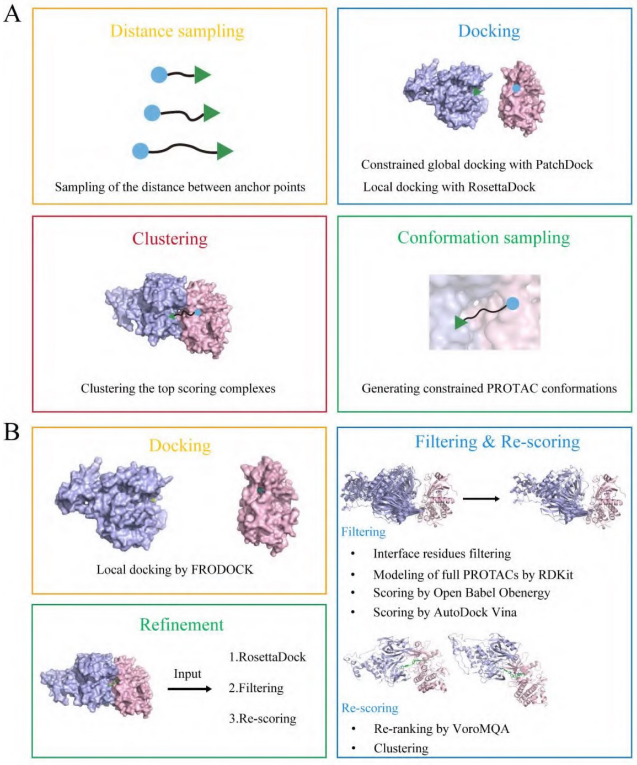

Figure 5. Modeling of PROTAC-mediated ternary complexes. (A) PRosettaC protocol; (B) PROTAC-Model protocol.

3.2.2. PROTAC-model

The PROTAC-model was proposed to integrate FRODOCK [66], RosettaDock [43], and several filtering and scoring methods (**Figure 5B**) [51]. Local protein-protein docking is performed using FRODOCK software [66], and the generated protein docking conformations are then screened. Firstly, docking conformations are filtered based on the number of interface residues. PROTAC conformations are then modeled using RDKit, eliminating unreasonable docking models. The energy of PROTAC conformations is calculated using the Open Babel Obenergy (Obenergy) [67] tool, and models with unfavorable atomic collisions are removed. AutoDock Vina (Vina) [68] is used to evaluate the binding mode of PROTAC to the protein-protein complex, and complexes with a Vina energy score less than 0 $kcalmol^{-1}$ are retained. The filtered models are further ranked by the VoroMQA method [69]. Finally, a clustering algorithm is applied to cluster these models. During the conformation screening process, the best model from each cluster provided by FRODOCK can also be selected for optimization in RosettaDock [43]. The generated models are then screened, reordered, and clustered similarly. This computational workflow integrates predicted ternary complex models with good degradation ability into the updated PROTAC-DB database.

3.3. Constructing PROTAC-target-E3-E2-ubiquitin complex structure model

Research indicates that the binding affinity of ternary complexes is not always correlated with target protein degradation. Although the formation of a target protein-PROTAC-E3 ligase ternary complex is key to achieving targeted degradation, the formation of the ternary complex does not always lead to TPD, suggesting that other factors can drive degradation [70]. Target protein ubiquitination is an important step following the formation of the ternary complex in the PROTAC-induced TPD process. PROTAC participates in a three-enzyme cascade reaction where E2/Ub binds to E3, transferring Ub to lysine residues on the target protein surface. Studies suggest that not every ternary complex conformation can be considered an “active“ conformation that can induce target ubiquitination, which may depend on the direction or distance from the accessible lysine on the target to Ub on E2. Bai *et al.* (2022) [64] proposed a structure-based computational method. By overlaying the E2-E3 ligase complex structure with the ternary complex, they constructed a ubiquitination model of the ternary complex bound to CRL4A (target/PROTAC/CRBN/DDB1/CUL4A/Rbx1/NEDD8/E2/Ub) using the Cullin-Ring ubiquitin ligase 4A (CRL4A) complex based on CRBN. The interactions between these proteins enable the transfer of Ub to the target protein (**Figure 6**) [64]. This method aims to investigate the relationship between ubiquitination ternary complex formation and lysine proximity in CRBN-based PROTAC systems from a structural modeling perspective.

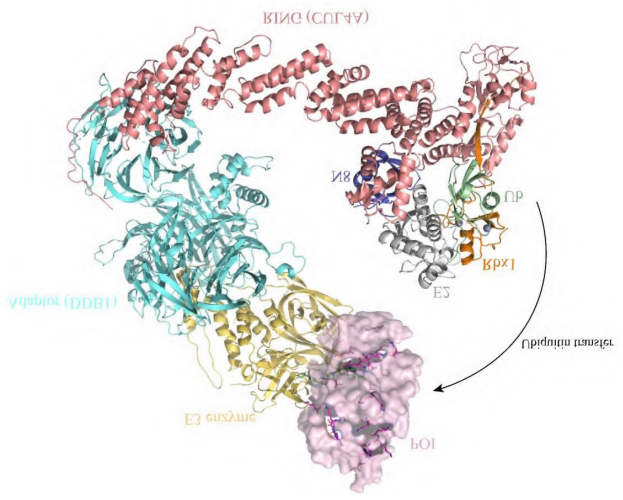

Figure 6. CRL4A ligase complex model.

4. Prospects of PROTAC technology

4.1. PTM-mediated protein stability techniques

Ubiquitination modification is an important type of ubiquitous protein post-translational modification (PTM). As a significant way of regulating protein function, PTM can regulate protein stability [71]. Ubiquitination regulates protein stability through the ubiquitin-proteasome system, degrading 80% of proteins in the body. Additionally, several common PTMs, such as methylation, phosphorylation, and acetylation, are also involved in regulating protein stability. These modifications are controlled by corresponding writer enzymes and eraser enzymes that regulate the modification level of substrate proteins. Currently, the number of detected PTMs far exceeds the number of functionally reported PTM sites, making it a significant challenge to interpret unknown functional PTM sites [72]. With the rapid development of precision medicine, research on modificomics is gradually deepening. For example, phosphoproteomics can comprehensively analyze phosphorylated proteins, enabling qualitative, localization, quantitative, and functional analysis of phosphorylation modifications [73]. Therefore, in studying PTM-driven protein stability regulation, functional PTM proteomics can be employed to systematically evaluate the impact of PTMs on protein stability.

The application of novel proteomics technologies is expected to greatly accelerate the identification and elucidation of PTM sites that regulate protein stability. Currently, quantitative proteomics based on high-throughput mass spectrometry (MS) and stable isotope labeling technology (SILAC) under cell culture conditions is primarily used to investigate which PTMs affect protein stability, enabling rapid screening of functional PTM sites. On the other hand, systematically measuring changes in protein stability after cellular perturbation can be achieved by interfering with the functions of PTM writer or eraser enzymes. Combining SILAC technology with tandem mass tags (TMT), a technique called multiplexed proteome dynamics profiling (mPDP), allows precise evaluation and comparison of the effects of multiple perturbations, revealing changes in protein stability under different perturbation conditions. Protein stability can be regulated by individual PTM sites or multiple PTMs. A single PTM acts as part of a PTM regulatory network formed through protein-protein interactions, where one PTM can serve as a promoter for the next PTM through network interactions [74]. The ways these PTMs participate in regulating protein stability suggest that the combined use of inhibitors targeting both eraser and writer enzymes to target PTM-modified proteins can be a potential drug design strategy [75]. Expanding the potential drug target space by targeting PTMs involved in the co-regulation of protein stability.

4.2. PTM-mediated proximity-inducing bimolecular techniques

In recent years, the research field of TPD has rapidly developed. With PROTAC entering clinical trials, drug development has entered the stage of heterobifunctional drugs. The PROTAC design concept has become a precursor to heterobifunctional molecule design strategies. This strategy utilizes proximity-inducing therapeutics, which connect recruitment molecules with target-binding molecules through rationally designed linkers, bringing the target protein close to functional proteins recruited by the recruitment molecules to achieve the desired biological effect [76]. The category of heterobifunctional molecules can be expanded based on the classification of recruited functional proteins. Heterobifunctional molecules can regulate the function of target proteins by recruiting their upstream functional proteins as effectors (such as E3 ubiquitin ligases, endosomes/lysosomes, ribonucleases L, protein phosphatases/kinases,

acetyltransferases, etc.), bringing effectors and target proteins closer to accelerate their interactions. Through this chemically induced proximity (CIP) technology, more and more heterobifunctional molecules are being designed to target undruggable proteins [10]. In targeted protein degradation, besides the ubiquitin-proteasome system, the lysosomal pathway is also an important protein degradation pathway, including autophagy-lysosome and endocytosis-lysosome pathways. Among them, techniques such as autophagy-targeting chimera (AUTAC) and autophagosome-tethering compounds (ATTEC) degrade target proteins through the autophagy-lysosome pathway; lysosome targeting chimera (LYTAC) technology degrades target proteins through the endocytosis-lysosome pathway (**Figure 7A**) [77]. These methods can selectively recognize proteins/organelles and transport them to lysosomes for degradation. The autophagy system plays a major role in lysosome-mediated intracellular material degradation, enabling the degradation of damaged organelles, intracellular debris, and other substrates. The endocytosis system mainly targets extracellular and membrane-associated proteins. Lysosome-based target protein degradation can overcome the limitations of proteasome-mediated degradation. However, the understanding and large-scale application of these methods are still in their infancy.

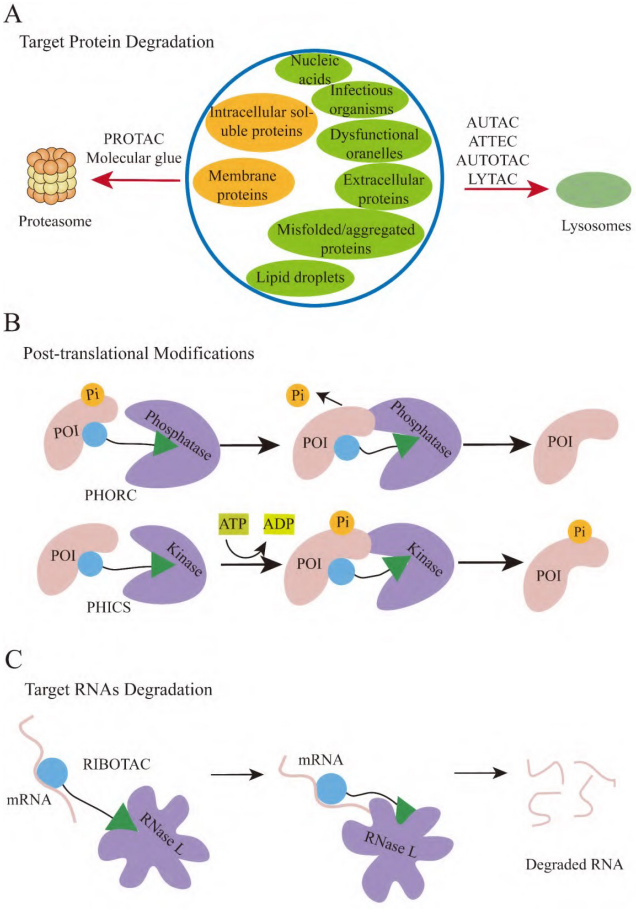

Figure 7. Different bifunctional molecules based on chemically induced proximity (CIP). (A) Target protein degradation; (B) Mechanism of PHORC and PHICS; (C): Target RNAs degradation.

In addition to targeted protein degradation, heterobifunctional molecules targeting PTMs can also alter protein function by modulating PTMs on target proteins. These heterobifunctional molecules recruit functional proteins that regulate PTMs, bringing them close to the target protein to exert their effects. Examples include phosphorylation-inducing chimera (PHICS) (**Fig 7B**), phosphatase-recruiting chimera (PHORC/phosphorylation targeting chimeras, PhosTAC) (**Fig 7B**), and acetylation-tagging-system (AceTAG) [76]. These heterobifunctional molecules precisely regulate the function of target proteins by altering their PTM status rather than their expression levels. Furthermore, beyond directly targeting protein PTMs, there are currently strategies to target pre-translational mechanisms, such as RNA degradation, RNA interference, and gene editing methods [71]. Ribonuclease Targeting Chimera (RIBOTAC) is a type of heterobifunctional molecule technology that degrades RNA [78], consisting of an RNA-targeting ligand, a recruiting ribonuclease (RNase L) ligand component, and a linker (**Figure 7C**). RIBOTAC recruits endogenous ribonuclease L to specific RNAs and activates ribonuclease, inducing proximity-based degradation of target RNAs [79]. However, the most significant limitation of this approach is its low cell permeability [11]. Additionally, designing highly selective RNA small molecule ligands is challenging, and prone to off-target effects, and RIBOTACs are not suitable for RNAs that function normally in the nucleus as they primarily act in the cytoplasm [80].

Heterobifunctional molecules are at the forefront of new trends in small-molecule drug design, recruiting different effectors to target proteins to regulate various biological processes. However, several critical issues regarding heterobifunctional molecules still need to be addressed, including hook effects, controllability, stability, and off-target effects. By developing heterobifunctional

molecules that recruit more endogenous effectors, the range of targeted regulatory mechanisms can be greatly enriched. Optimizing heterobifunctional molecules to overcome their drug ability is a significant research challenge in the coming years [76].

4.3. Application of bioinformatics and computational simulation techniques in PROTAC development

Among the 600+ E3 ligases in the human genome, less than 2% have been involved in targeted protein degradation studies. While systematic analyses of target proteins have been conducted to evaluate their PROTAC ability [37], the E3 ligase space has not been systematically quantified. Recruiting more potential E3 ligases is crucial for achieving precise targeted protein degradation. Beyond the strategies described in Section 2.1 to recruit more potential E3 ligases based on expression profiles, structural availability, and functional necessity, recent efforts by Liu *et al.* (2023) [22] have integrated data from Ge *et al.* (2018) [81], Ubihub [20], and UbiBrowser 2.0 [82] to obtain a comprehensive list of E3 ligases. They further propose expanding the E3 ligase space based on chemical ligands, protein-protein interactions (PPI), PPI interface analysis, and cellular location of E3 ligases. This approach involves collecting E3 ligase ligand data through database searches, predicting interactions between E3 ligases and drugs using deep learning-based virtual screening models, constructing comprehensive PPI maps based on interactions between E3 ligases and target proteins, analyzing PPI interface information to identify mutations that may disrupt these interactions, and determining the cellular location of E3 ligases to discover more potential novel E3 ligases. Despite current limitations, recruiting more potential E3 ligases is significant for expanding the target protein space, enhancing selective degradation, and guiding PROTAC design.

Regarding the construction of PROTAC ternary complex structural models, while several modeling methods have been reported in recent years, the binding of ternary complexes cannot be described by solely calculating the interaction energy between PROTAC and each protein [40]. The binding stability and synergism of PROTAC-mediated ternary complexes are key factors determining PROTAC degradation efficiency. The emergence of hook effects is detrimental to the stable existence of ternary complexes, and higher synergism is believed to be associated with weaker hook effects. Therefore, predicting synergism is crucial for PROTAC design and optimization. Li *et al.* (2022) [40] obtained low-energy structures of PROTAC ternary complexes through molecular dynamics simulations, calculated binding energies and predicted synergism using MM/GBSA, and improved the accuracy of ternary complex conformation prediction by introducing a new scoring and ranking system based on PRosettaC.

On the other hand, with the booming development of artificial intelligence methods, increasing studies are combining deep learning approaches to predict the structural models of ternary complexes, predict PROTAC degradation efficiency, and guide PROTAC design. Zheng *et al.* (2022) [83] proposed a deep generative model (PROTAC-RL) to design and sample PROTACs for given target proteins and E3 ligands. They used reinforcement learning to generate PROTACs with optimal pharmacokinetic properties and employed PRosettaC [49] and molecular dynamics simulations for conformation clustering and screening analysis. This combination of deep learning and molecular simulation can facilitate rational PROTAC design and optimization. Due to the limitations of current ternary complex modeling methods, Li *et al.* (2022) [65] circumvented the modeling process by extracting five components from determined protein-ligand structures: target protein pocket, E3 ligase pocket, a target protein-ligand, E3 ligase ligand, and linker. They then used five modules to extract corresponding features and proposed a deep learning model called DeepPROTACs based on graph neural networks to predict the degradation efficacy of designed PROTACs for target proteins.

Recent research [15] has introduced a new method for generating PROTAC ternary complex structural models. Considering the rationality of protein conformations and the geometry of the linker itself, the ligand-generated linker is split into two parts and connected to the E3 ligase-ligand complex and the target protein-ligand complex, respectively. Then, based on the Fourier fast sampling algorithm, protein-protein complex conformations are generated to obtain low-energy conformations of target proteins and E3 ligases, along

with potentially reasonable linkers. These are further optimized through energy minimization, clustering, and ranking. By testing published ternary complex structures, this model can predict ternary complex structures with high precision and predict synergism and degradation efficiency of ternary complexes based on the method proposed by Bai *et al.* (2021) [50]. This approach integrates protein structures into deep learning models, simultaneously obtaining ternary complex structures and PROTAC degradation efficiency. The combination of bioinformatics, computational simulation, and AI holds significant promise in predicting ternary complex formation, degradation efficiency prediction, and guiding PROTAC design.

5. Summary

PROTAC technology utilizes the ubiquitin-proteasome system for the ubiquitination and degradation of target proteins. Compared to other small molecule drugs and macromolecular drugs, PROTAC exhibits unique advantages. However, this technology also faces numerous challenges, particularly in improving selectivity and specificity, which largely depends on the choice of E3 ubiquitin ligases. Currently, there are very few E3 ligases that can be used, so expanding the space of E3 ubiquitin ligases is crucial for the rational design of PROTACs to enhance selectivity. Additionally, PROTACs do not need to bind tightly to proteins, the key to achieving degradation is the formation of a stable ternary complex. Efficient and selective degradation relies, to some extent, on the plasticity of PROTAC-induced low-energy binding conformations and inter-protein contacts. Therefore, structurally studying the formation of ternary complexes is significant for understanding selective degradation. Nevertheless, there are currently few reported crystal structures of ternary complexes, and molecular simulation can help understand how PROTACs induce the interaction between target proteins and E3 ligases to form ternary complexes, guiding PROTAC design and deepening the understanding of PROTAC structure-activity relationships. With the rapid development of bioinformatics and computational simulation methods, many studies have constructed ternary complex models by combining deep learning and structural simulation to address the limited number of ternary complex crystal structures and the accuracy of modeling methods. These studies have predicted PROTAC degradation efficiency, positively contributing to guiding PROTAC design. After years of research and development, the PROTAC field has spawned many new technologies based on chemically induced proximity (CIP) theory, and the field of drug design based on CIP theory has entered the stage of heterodimeric bifunctional molecules. Overall, PROTAC's future prospects in drug discovery are vast and will open up new possibilities for expanding the druggable proteome.

Funding

National Natural Science Foundation of China (Project No.: 22377089); Key Discipline Construction Project of Jiangsu Higher Education Institutions (PAPD); College Student Innovation and Entrepreneurship Training Program (Project No.: 202210285179Y)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Sakamoto KM, Kim KB, Kumagai A, et al., 2001, PROTACs: Chimeric Molecules That Target Proteins to the Skp1-Cullin-F Box Complex for Ubiquitination and Degradation. Proceedings of the National Academy of Sciences of the

United States of America, 98: 8554–8559.

[2] Sakamoto KM, Kim KB, Verma R, et al., 2003, Development of PROTACs to Target Cancer-Promoting Proteins for Ubiquitination and Degradation. Molecular & Cellular Proteomics, 2: 1350–1358.

[3] Yao T, Xiao H, Wang H, et al., 2022, Recent Advances in PROTACs for Drug-Targeted Protein Research. International Journal of Molecular Sciences, 23: 10328.

[4] Sheng XY, Wu SH, Li BL, et al., 2021, Advances in the Optimization of the Linker in Proteolysis-Targeting Chimeras (PROTAC). Acta Pharmacologica Sinica, 56: 445–455.

[5] Xie BH, Hu ZY, Ning WT, et al., 2020, The Research Progress of PROTACs for Breast Cancer Treatment. Acta Pharmacologica Sinica, 55: 2053–2061.

[6] Zeng S, Huang W, Zheng X, et al., (2021), Proteolysis Targeting Chimera (PROTAC) in Drug Discovery Paradigm: Recent Progress and Future Challenges. European Journal of Medicinal Chemistry, 210: 112981.

[7] Xu YR, Zhang QS, Wu JY, et al., 2021, Recent Progress in Targeting Degradation of FAK Based on PROTAC. Acta Pharm Sin, 56: 1571–1579.

[8] Maneiro MA, Forte N, Shchepinova MM, et al., 2020, Antibody-PROTAC Conjugates Enable HER2-Dependent Targeted Protein Degradation of BRD4. ACS Chemical Biology, 15: 1306–1312.

[9] Imaide S, Riching KM, Makukhin N, et al., 2021, Trivalent PROTACs Enhance Protein Degradation via Combined Avidity and Cooperativity. Nature Chemical Biology, 17: 1157–1167.

[10] Li K, Crews CM, 2022, PROTACs: Past, Present, and Future. Chemical Society Reviews, 51: 5214–5236.

[11] Wang C, Zheng C, Wang H, et al., 2022, The State of the Art of PROTAC Technologies for Drug Discovery. European Journal of Medicinal Chemistry, 235: 114290.

[12] Guenette RG, Yang SW, Min J, et al., 2022, Target and Tissue Selectivity of PROTAC Degraders. Chem Soc Rev, 51: 5740–5756.

[13] Ward JA, Perez-Lopez C, Mayor-Ruiz C, 2023, Biophysical and Computational Approaches to Study Ternary Complexes: A 'Cooperative Relationship' to Rationalize Targeted Protein Degradation. ChemBioChem, 24: e202300163.

[14] Gadd MS, Testa A, Lucas X, et al., 2017, Structural Basis of PROTAC Cooperative Recognition for Selective Protein Degradation. Nature Chemical Biology, 13: 514–521.

[15] Ignatov M, Jindal A, Kotelnikov S, et al., 2023, High Accuracy Prediction of PROTAC Complex Structures. Journal of the American Chemical Society, 145: 7123–7135.

[16] Nowak RP, Deangelo SL, Buckley D, et al., 2018, Plasticity in Binding Confers Selectivity in Ligand-Induced Protein Degradation. Nature Chemical Biology, 14: 706–714.

[17] Ramachandran S, Ciulli A, 2021, Building Ubiquitination Machineries: E3 Ligase Multi-Subunit Assembly and Substrate Targeting by PROTACs and Molecular Glues. Current Opinion in Structural Biology, 67: 110–119.

[18] Bernassola F, Chillemi G, Melino G, 2019, HECT-Type E3 Ubiquitin Ligases in Cancer. Trends in Biochemical Sciences, 44: 1057–1075.

[19] Wang P, Dai X, Jiang W, et al., 2020, RBR E3 Ubiquitin Ligases in Tumorigenesis. Semin Cancer Biol, 67: 131–144.

[20] Liu L, Damerell DR, Koukouflis L, et al., 2019, UbiHub: A Data Hub for the Explorers of Ubiquitination Pathways. Bioinformatics, 35: 2882–2884.

[21] Bekes M, Langley DR, Crews CM, 2022, PROTAC Targeted Protein Degraders: The Past Is Prologue. Nature Reviews Drug Discovery, 21: 181–200.

[22] Liu Y, Yang J, Wang T, et al., 2023, Expanding PROTACtable Genome Universe of E3 Ligases. Nature Communications, 14: 6509.

[23] Weng G, Shen C, Cao D, et al., 2021, PROTAC-DB: An Online Database of PROTACs. Nucleic Acids Research, 49:

D1381–D1387.

[24] Ward CC, Kleinman JI, Brittain SM, et al., 2019, Covalent Ligand Screening Uncovers a RNF4 E3 Ligase Recruiter for Targeted Protein Degradation Applications. ACS Chemical Biology, 14: 2430–2440.

[25] Ishida T, Ciulli A, 2021, E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. SLAS Discovery, 26: 484–502.

[26] Tunyasuvunakool K, Adler J, Wu Z, et al., 2021, Highly Accurate Protein Structure Prediction for the Human Proteome. Nature, 596: 590–596.

[27] Baek M, Dimaio F, Anishchenko I, et al., 2021, Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. Science, 2021, 373: 871–876.

[28] Kannt A, Dikic I, 2021, Expanding the Arsenal of E3 Ubiquitin Ligases for Proximity-Induced Protein Degradation. Cell Chemical Biology, 28: 1014–1031.

[29] Ehrlich KC, Baribault C, Ehrlich M, 2020, Epigenetics of Muscle- and Brain-Specific Expression of KLHL Family Genes. International Journal of Molecular Sciences, 21: 8394.

[30] Liu QY, Lei JX, Sikorska M, et al., 2008, A Novel Brain-Enriched E3 Ubiquitin Ligase RNF182 Is Up-Regulated in the Brains of Alzheimer's Patients and Targets ATP6V0C for Degradation. Molecular Neurodegeneration, 2008, 3: 4.

[31] Khan S, Zhang X, Lv D, et al., 2019, A Selective BCL-XL PROTAC Degrader Achieves Safe and Potent Antitumor Activity. Nature Medicine, 25: 1938–1947.

[32] Pacini C, Dempster JM, Boyle I, et al., 2021, Integrated Cross-Study Datasets of Genetic Dependencies in Cancer. Nature Communication, 12: 1661.

[33] Meyers RM, Bryan JG, McFarland JM, et al., 2017, Computational Correction of Copy Number Effect Improves Specificity of CRISPR-Cas9 Essentiality Screens in Cancer Cells. Nature Genetics, 49: 1779–1783.

[34] Nieto-Jimenez C, Morafraile EC, Alonso-Moreno C, et al., 2022, Clinical Considerations for the Design of PROTACs in Cancer. Molecular Cancer, 21: 67.

[35] Weon JL, Potts PR, 2015, The MAGE Protein Family and Cancer. Current Opinion in Cell Biology, 37: 1–8.

[36] Feng Y, Gao J, Yang M, 2011, When MAGE Meets RING: Insights Into Biological Functions of MAGE Proteins. Protein Cell, 2: 7–12.

[37] Schneider M, Radoux CJ, Hercules A, et al., 2021, The PROTACtable Genome. Nature Reviews Drug Discovery, 20: 789–797.

[38] Kim W, Bennett EJ, Huttlin EL, et al., 2011, Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. Molecular Cell, 44: 325–340.

[39] Mathieson T, Franken H, Kosinski J, et al., 2018, Systematic Analysis of Protein Turnover in Primary Cells. Nat Commun, 9: 689.

[40] Li W, Zhang J, Guo L, et al., 2022, Importance of Three-Body Problems and Protein-Protein Interactions in Proteolysis-Targeting Chimera Modeling: Insights From Molecular Dynamics Simulations. J Chem Inf Model, 62: 523–532.

[41] Schiedel M, Herp D, Hammelmann S, et al., 2018, Chemically Induced Degradation of Sirtuin 2 (Sirt2) by a Proteolysis Targeting Chimera (PROTAC) Based on Sirtuin Rearranging Ligands (SirReals). Journal of Medicinal Chemistry, 61: 482–491.

[42] Dominguez C, Boelens R, Bonvin AM, 2003, HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. Journal of the American Chemical Society, 125: 1731–1737.

[43] Gray JJ, Moughon S, Wang C, et al., 2003, Protein-Protein Docking With Simultaneous Optimization of Rigid-Body Displacement and Side-Chain Conformations. Journal of Molecular Biology, 331: 281–299.

[44] Pierce BG, Wiehe K, Hwang H, et al., 2014, ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes

and Symmetric Multimers. Bioinformatics, 30: 1771–1773.

[45] Yan Y, Zhang D, Zhou P, et al., 2017, HDOCK: A Web Server for Protein-Protein and Protein-DNA/RNA Docking Based on a Hybrid Strategy. Nucleic Acids Research, 45: W365–W373.

[46] Kozakov D, Hall DR, Xia B, et al., 2017, The ClusPro Web Server for Protein-Protein Docking. Nature Protocols, 12: 255–278.

[47] Schneidman-Duhovny D, Inbar Y, Nussinov R, et al., 2005, PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. Nucleic Acids Research, 33: W363–W367.

[48] Jimenez-Garcia B, Roel-Touris J, Romero-Durana M, et al., 2018, LightDock: A New Multi-Scale Approach to Protein-Protein Docking. Bioinformatics, 34: 49–55.

[49] Zaidman D, Prilusky J, London N, 2020, PRosettaC: Rosetta-Based Modeling of PROTAC-Mediated Ternary Complexes. Journal of Chemical Information and Modeling, 60: 4894–4903.

[50] Bai N, Miller SA, Andrianov GV, et al., 2021, Rationalizing PROTAC-Mediated Ternary Complex Formation Using Rosetta. Journal of Chemical Information and Modeling, 61: 1368–1382.

[51] Weng G, Li D, Kang Y, et al., 2021, Integrative Modeling of PROTAC-Mediated Ternary Complexes. Journal of Medicinal Chemistry, 64: 16271–16281.

[52] Hawkins PC, Skillman AG, Warren GL, et al., 2010, Conformer Generation with OMEGA: Algorithm and Validation Using High-Quality Structures from the Protein Data Bank and Cambridge Structural Database. Journal of Chemical Information and Modeling, 50: 572–584.

[53] Saikia S, Bordoloi M, 2019, Molecular Docking: Challenges, Advances, and Its Use in Drug Discovery Perspective. Current Drug Targets, 20: 501–521.

[54] Salmaso V, Moro S, 2018, Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. Frontiers in Pharmacology, 9: 923.

[55] Hollingsworth SA, Dror RO, 2018, Molecular Dynamics Simulation for All. Neuron, 99: 1129–1143.

[56] Weerakoon D, Carbajo RJ, De Maria L, et al., 2022, Impact of PROTAC Linker Plasticity on the Solution Conformations and Dissociation of the Ternary Complex. Journal of Chemical Information and Modeling, 62: 340–349.

[57] Liao J, Nie X, Unarta IC, et al., 2022, *In Silico* Modeling and Scoring of PROTAC-Mediated Ternary Complex Poses. Journal of Medicinal Chemistry, 65: 6116–6132.

[58] Skjaerven L, Martinez A, Reuter N, 2011, Principal Component and Normal Mode Analysis of Proteins; A Quantitative Comparison Using the GroEL Subunit. Proteins, 79: 232–243.

[59] Bauer JA, Pavlovic J, Bauerova-Hlinkova V, 2019, Normal Mode Analysis as a Routine Part of a Structural Investigation. Molecules, 24: 3293.

[60] Alexandrov V, Lehnert U, Echols N, et al., 2005, Normal Modes for Predicting Protein Motions: A Comprehensive Database Assessment and Associated Web Tool. Protein Science, 14: 633–643.

[61] Drummond ML, Williams CI, 2019, In Silico Modeling of PROTAC-Mediated Ternary Complexes: Validation and Application. Journal of Chemical Information and Modeling, 59: 1634–1644.

[62] Bondeson DP, Smith BE, Burslem GM, et al., 2018, Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead. Cell Chemical Biology, 25: 78–87.

[63] Smith BE, Wang SL, Jaime-Figueroa S, et al., 2019, Differential PROTAC Substrate Specificity Dictated by Orientation of Recruited E3 Ligase. Nature Communications, 10: 131.

[64] Bai N, Riching KM, Makaju A, et al., 2022, Modeling the CRL4A Ligase Complex to Predict Target Protein Ubiquitination Induced by Cereblon-Recruiting PROTACs. Journal of Biological Chemistry, 298: 101653.

[65] Li F, Hu Q, Zhang X, et al., 2022, DeepPROTACs is a Deep Learning-Based Targeted Degradation Predictor for

PROTACs. Nature Communications, 13: 7133.

[66] Garzon JI, Lopez-Blanco JR, Pons C, et al., 2009, FRODOCK: A New Approach for Fast Rotational Protein-Protein Docking. Bioinformatics, 25: 2544–2551.

[67] O'Boyle NM, Banck M, James CA, et al., 2011, Open Babel: An Open Chemical Toolbox. Journal of Cheminformatics, 3: 33.

[68] Trott O, Olson AJ, 2010, AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. Journal of Computational Chemistry, 31: 455–461.

[69] Olechnovic K, Venclovas C, 2017, VoroMQA: Assessment of Protein Structure Quality Using Interatomic Contact Areas. Proteins, 85: 1131–1145.

[70] Han B, 2020, A Suite of Mathematical Solutions to Describe Ternary Complex Formation and Their Application to Targeted Protein Degradation by Heterobifunctional Ligands. Journal of Biological Chemistry, 295: 15280–15291.

[71] Amirian R, Azadi Badrbani M, Izadi Z, et al., 2023, Targeted Protein Modification as a Paradigm Shift in Drug Discovery. European Journal of Medicinal Chemistry, 260: 115765.

[72] Liang Z, Liu T, Li Q, et al., 2023, Deciphering the Functional Landscape of Phosphosites with Deep Neural Network. Cell Reports, 42: 113048.

[73] Zhu F, Yang S, Meng F, et al., 2022, Leveraging Protein Dynamics to Identify Functional Phosphorylation Sites Using Deep Learning Models. Journal of Chemical Information and Modeling, 62: 3331–3345.

[74] Zhu F, Deng L, Dai Y, et al., 2023, PPICT: An Integrated Deep Neural Network for Predicting Inter-Protein PTM Cross-Talk. Briefings in Bioinformatics, 24: bbad052.

[75] Lee JM, Hammaren HM, Savitski MM, et al., 2023, Control of Protein Stability by Post-Translational Modifications. Nature Communications, 14: 201.

[76] Hua L, Zhang Q, Zhu X, et al., 2022, Beyond Proteolysis-Targeting Chimeric Molecules: Designing Heterobifunctional Molecules Based on Functional Effectors. Journal of Medicinal Chemistry, 65: 8091–8112.

[77] Paudel RR, Lu D, Roy Chowdhury S, et al., 2023, Targeted Protein Degradation via Lysosomes. Biochemistry, 62: 564–579.

[78] Borgelt L, Haacke N, Lampe P, et al., 2022, Small-Molecule Screening of Ribonuclease L Binders for RNA Degradation. Biomedicine & Pharmacotherapy, 154: 113589.

[79] Meyer SM, Tanaka T, Zanon PRA, et al., 2022, DNA-Encoded Library Screening to Inform Design of a Ribonuclease Targeting Chimera (RiboTAC). Journal of the American Chemical Society, 144: 21096–21102.

[80] Dey SK, Jaffrey SR, 2019, RIBOTACs: Small Molecules Target RNA for Degradation. Cell Chemical Biology, 26: 1047–1049.

[81] Ge Z, Leighton JS, Wang Y, et al., 2018, Integrated Genomic Analysis of the Ubiquitin Pathway Across Cancer Types. Cell Reports, 23: 213–226.

[82] Wang X, Li Y, He M, et al., 2022, UbiBrowser 2.0: A Comprehensive Resource for Proteome-Wide Known and Predicted Ubiquitin Ligase/Deubiquitinase-Substrate Interactions in Eukaryotic Species. Nucleic Acids Research, 50: D719–D728.

[83] Zheng SJ, Tan YH, Wang ZY, et al., 2022, Accelerated Rational PROTAC Design via Deep Learning and Molecular Simulations. Nature Machine Intelligence, 4: 739–748.