Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

GENERATIVE MODELS FOR PROTAC LINKER DESIGN USING TERNARY COMPLEX GEOMETRY AND DEGRADATION DATA

Ethan Wright[1], Chloe Bennett[1]*, Jack Turner[2]

1. ***Department of Pharmaceutical AI Analytics, Faculty of Pharmacy, University of Leeds, Leeds, United Kingdom.***
2. ***Department of Computational Drug Systems, Faculty of Medicine, University of Sheffield, Sheffield, United Kingdom.***

ARTICLE INFO

Received:
11 March 2026
Received in revised form:
20 June 2026
Accepted:
22 June 2026
Available online:
28 June 2026

Keywords: PROTACs, Generative chemistry, Linker design, Ternary complex, Targeted protein degradation, Diffusion models

ABSTRACT

PROTACs are bifunctional therapeutic agents that achieve selective protein degradation by recruiting a target protein to an E3 ligase, with the linker connecting the two ligands serving as a critical determinant of ternary complex formation, cellular activity, and degradation efficacy. Current linker design is largely empirical, relying on iterative synthesis of chemically familiar linkers, which constrains the exploration of broader chemical space and slows the discovery of PROTACs with optimal geometry and degradation profiles. To address this, we propose a generative chemistry framework for PROTAC linker design that leverages ternary complex structural information and degradation activity annotations to produce linkers compatible with both molecular geometry and pharmacological intent. Conditional generative models, such as diffusion models or variational autoencoders, can be trained on PROTAC structures paired with degradation metrics like DC50 and Dmax, and conditioned during generation on the protein of interest, E3 ligase, warhead structures, ternary complex geometry, and desired degradation profile. The resulting linkers are expected to be chemically valid, compatible with the two warheads, and consistent with stereochemical constraints derived from ternary complex models, preserving key binding interactions while enabling productive POI–E3 proximity. By moving beyond empirical variation, structure-informed and degradation-aware generative design could accelerate PROTAC development and enable systematic exploration of linker space for targeted protein degradation.

To Cite This Article: Wright E, Bennett C, Turner J. Generative Models for PROTAC Linker Design Using Ternary Complex Geometry and Degradation Data. Pharmacophore. 2026;17(3):123-31. https://doi.org/10.51847/JR4VTce12r

Introduction

PROTACs induce targeted protein degradation by recruiting an E3 ligase to a protein of interest, thereby forming a ternary complex that can trigger ubiquitination and proteasomal degradation. The linker connecting the POI-binding ligand and E3-ligase ligand is not a passive spacer; it shapes the spatial relationship between the two proteins and can determine whether a productive complex forms [1]. Structural analyses of cooperative PROTAC complexes have shown that productive degradation depends on the induced protein–protein interface as well as on how the linker positions the two warheads [2]. As a result, linker optimization remains one of the central design challenges in PROTAC discovery, because small changes in length, composition, or attachment site can alter ternary complex geometry and degradation behavior [3].

Traditional PROTAC campaigns often explore PEG, alkyl, or mixed linkers through empirical structure–activity relationships. Reviews of linker design emphasize that linker length, polarity, rigidity, and conformational plasticity can influence not only degradation potency but also permeability and ternary complex stability [4]. Computational studies of linker plasticity further suggest that linkers can modulate solution conformations and ternary complex dissociation, making simple distance-based design rules insufficient. Because the accessible linker chemical space is much larger than the set typically synthesized, empirical linker variation can leave productive geometries unexplored.

Generative chemistry offers a route to systematic linker exploration, especially when models are conditioned on structural constraints. General 3D linker design methods have demonstrated that molecular linkers can be generated under geometric constraints between fragments [5], and pharmacophore-aware generative approaches suggest that spatial constraints can be incorporated directly into design objectives [6]. PROTAC-specific generative work has begun to adapt reinforcement learning

Corresponding Author: Chloe Bennett; Department of Pharmaceutical AI Analytics, Faculty of Pharmacy, University of Leeds, Leeds, United Kingdom. E-mail: chloe.bennett@gmail.com.

and 3D-aware generation to bifunctional molecules [7], while broader linker-generation frameworks show how synthetic and structural objectives can guide molecule construction [8]. However, PROTACs remain uniquely difficult because the generated linker must connect two large functional ligands while supporting a protein–protein interaction rather than merely optimizing isolated ligand affinity.

The thesis of this EAI article is that a PROTAC linker generator should jointly learn ternary complex geometry and degradation activity rather than treating linker design as an unconstrained molecular-generation task. Public resources such as PROTAC-DB provide curated PROTAC structures and activity annotations that can support data-driven model development [9], while updated database releases expand the scope of degradation-associated information available for computational modeling [10, 11]. Structure-based tools such as PRosettaC and Rosetta-based ternary complex modeling provide complementary geometric information that can condition or evaluate generated linkers [12, 13]. By integrating structural compatibility, linker chemistry, and degradation labels into a single design framework, a generative model could propose linkers that are not only novel but also mechanistically plausible.

Background

PROTAC Mechanism and Linker Design Principles

PROTAC mechanism depends on the formation of a ternary complex in which the target protein and E3 ligase are brought into a productive orientation. Linker length and composition can shift this orientation by controlling the distance, flexibility, and conformational ensemble available to the two warheads [3]. Medicinal chemistry analyses have emphasized that rigid, semi-rigid, and flexible linkers may each be useful depending on the POI–E3 pair and the required interface geometry [4]. Studies of linker-dependent folding and permeability also indicate that linker choice affects properties beyond ternary complex formation, including conformational shielding and cellular exposure.

Structural Biology of Ternary Complexes

Structural biology has provided direct evidence that PROTAC activity can arise from cooperative recognition rather than simple binary affinity. The crystal structure of a VHL-based PROTAC ternary complex showed how the degrader can stabilize new contacts between the target and E3 ligase [1], while later work highlighted plasticity in ligand-induced degradation complexes [2]. Computational modeling approaches, including Rosetta-based sampling and integrative ternary complex construction, have been developed to extend these insights to systems lacking solved structures [12, 14]. Docking- and pose-scoring strategies further support the use of modeled ternary complexes as design constraints for linker generation [15].

Degradation Activity Data and Datasets

Experimental degradation data include endpoints such as DC50, Dmax, and degradation kinetics, each capturing a different aspect of cellular degrader performance. PROTAC-DB was introduced as an online database of PROTACs to organize chemical structures, targets, E3 ligases, and activity annotations for computational use [9]. Subsequent versions expanded the database and improved its utility for model development by adding broader pharmacological and pharmacokinetic information [10, 11]. These resources remain heterogeneous, however, because degradation assays differ across cell lines, time points, target abundance, and readout technologies, making harmonization essential before training predictive or generative models.

Generative Models for Drug Design and Bifunctional Molecules

Generative models for linker design include variational autoencoders, reinforcement learning systems, and diffusion-based models that can construct molecules under spatial or pharmacological constraints. Deep generative linker models have shown how two molecular fragments can be joined while respecting 3D geometry [5], and Link-INVENT demonstrated reinforcement learning for linker generation with optimization objectives [8]. Diffusion-based linker design offers another route because it can sample molecular structures conditioned on fragment coordinates and spatial context [16]. PROTACs extend these challenges because the generated linker must preserve two warhead attachment points, tolerate large molecular size and flexibility, and promote a productive ternary complex rather than merely joining fragments.

Computational Evaluation of PROTACs

Computational PROTAC evaluation commonly combines ternary complex modeling, conformational analysis, and activity prediction. PRosettaC and related Rosetta-based approaches use structural sampling to estimate plausible POI–PROTAC–E3 assemblies [12, 13], while integrative modeling and pose-scoring workflows evaluate whether a candidate linker can support a favorable ternary orientation [14, 15]. Molecular dynamics and conformational analyses can further assess linker plasticity, solution preferences, and the likelihood that the PROTAC can access complex-compatible conformations. Machine learning degradation predictors such as DeepPROTACs offer complementary activity-focused scoring that can be used as a surrogate evaluator during or after generation [17].

Model Development Overview

High-Level Generation Pipeline

In the proposed pipeline, a user specifies the target protein, E3 ligase, and warhead structures, and may also specify a desired degradation profile. The model then generates linker candidates as SMILES strings, molecular graphs, or 3D linker

conformations, following the design logic of PROTAC-oriented reinforcement learning and 3D linker generation frameworks [7, 8]. A scoring module evaluates whether each linker is chemically valid, compatible with the warhead attachment sites, and plausible in a modeled ternary complex, drawing on structure-based evaluation strategies developed for PROTAC complexes [12, 15]. Candidates that pass these filters would be prioritized for synthesis planning and biological testing rather than presented as validated degraders.

Core Inputs and Outputs

The core inputs include POI identity, E3 ligase identity, warhead structures, linker attachment points, available ternary complex geometry, and degradation activity labels. PROTAC databases provide structured examples of target, E3 ligase, compound, and activity relationships that can form the foundation of training data [9, 10]. Degradation-prediction models such as DeepPROTACs show how chemical and biological features can be mapped to degradation outcomes, suggesting how activity labels can be integrated into generative conditioning [17]. The outputs are proposed linker candidates, together with model-derived annotations such as predicted degradation tendency, structural compatibility, and confidence estimates that should guide prioritization rather than replace experiments.

Design Principles

The generated linker must be chemically valid, synthetically plausible, and compatible with the two warhead attachment chemistries. It should preserve binary binding interactions while positioning the POI and E3 ligase in a ternary geometry that is expected to support productive ubiquitination, consistent with structural studies of cooperative PROTAC recognition [1, 2]. Linker designs should also account for conformational plasticity, because flexible PROTACs may adopt ensembles that differ between solution, binary binding, and ternary complex states. These principles make PROTAC linker generation a constrained molecular-design problem rather than a general de novo molecule-generation task.

Data Sources and Input Representations

Curation of PROTAC Training Data

Training data would be curated from PROTAC-DB releases and peer-reviewed literature, including compound structures, POI identities, E3 ligases, linker definitions, warhead annotations, and degradation endpoints. The original PROTAC-DB established an organized public resource for PROTAC chemical and bioactivity data [9], while later versions expanded its coverage and added information useful for model development [10, 11]. Activity fields such as DC50 and Dmax should be standardized conceptually across assay contexts, with metadata retained for target, cell type, incubation time, and detection method. Because degradation datasets remain heterogeneous, curation should emphasize traceability and avoid collapsing distinct biological conditions into a single unqualified potency label.

Representing Warheads and Linkers

The POI-binding and E3-binding warheads can be represented as molecular graphs, fingerprints, or learned embeddings, while the linker should be encoded with explicit attachment points. PROTAC-Splitter illustrates the importance of automatically identifying PROTAC substructures, because generative models must distinguish the two warheads from the linker before learning meaningful design rules [18]. For molecular generation, graph-based and SMILES-based representations can encode atom types and connectivity, whereas 3D representations can capture the spatial constraints needed for linker geometry [5, 16]. A PROTAC-aware representation should also respect the asymmetry of the two ligand ends, because reversing warhead context may change the biological interpretation of an otherwise similar linker.

Incorporating Ternary Complex Information

For training examples with solved or modeled ternary complexes, geometric descriptors can be extracted from the relationship between the two warhead attachment points and the surrounding POI–E3 interface. These descriptors may include attachment-point separation, relative orientation, linker exit vectors, steric clearance, and interface proximity, following the logic of structure-based PROTAC modeling [12, 14]. Pose-scoring methods for PROTAC-mediated ternary complexes can help identify whether a generated linker is likely to support a plausible assembly [15]. Such geometry can be used either as conditioning information during generation or as an auxiliary target that encourages the model to learn structural compatibility rather than merely reproduce known linkers.

Table 1 defines the representational logic required to convert PROTAC linker design constraints into model-readable inputs that preserve geometry, chemistry, degradation context, and experimental interpretability.

Table 1. Constraint-to-Representation Framework for Generative PROTAC Linker Design

Design constraint	Required representation	Why it matters for linker generation	Model-use function	Failure risk if omitted
POI-binding warhead identity	Molecular graph, fingerprint, or learned warhead embedding	Defines the target-facing end of the bifunctional molecule and preserves the biological context of the POI interaction	Conditions the generator on the POI-side ligand environment	Generated linkers may disrupt POI binding or ignore target-specific exit-vector constraints

E3-ligase warhead identity	Molecular graph, fingerprint, or learned E3-warhead embedding	Defines the ligase-facing end and captures ligase-specific linker tolerance	Conditions linker generation on the E3-recruitment context	Designs may appear chemically valid but fail to support productive E3 engagement
Linker attachment points	Explicit atom-level attachment labels and exit-vector encoding	Determines where the linker can be connected without changing warhead function	Constrains graph construction and virtual warhead attachment	Invalid or chemically unrealistic PROTACs may be produced
Attachment-point distance	3D distance between POI-warhead and E3-warhead connection atoms	Provides a first-order geometric constraint on linker span	Guides spatial linker length and conformer sampling	Linkers may be too short, too long, or geometrically strained
Relative warhead orientation	3D vector or rotational descriptor between ligand exit directions	Captures whether the linker can support the required ternary orientation	Conditions geometry-aware generation	A linker may connect fragments but force an unproductive POI–E3 arrangement
Ternary complex interface context	POI–E3 contact region, steric clearance, and proximity descriptors	Indicates whether linker placement is compatible with induced protein–protein interaction	Penalizes steric conflict and supports productive complex geometry	Generated molecules may interfere with the interface needed for ubiquitination
Linker chemistry	Atom types, bond types, rigidity, polarity, heteroatom pattern, ring content	Determines flexibility, solubility, permeability, and conformational behavior	Defines the chemical search space of the generator	The model may overproduce familiar or unsuitable PEG/alkyl-like linkers
Degradation potency	Harmonized DC50 label with assay context	Provides a potency-oriented activity signal	Conditions or ranks candidates toward lower effective degradation concentration	The model may optimize geometry without pharmacological relevance
Degradation extent	Harmonized Dmax label with assay context	Captures maximal achievable degradation rather than potency alone	Helps distinguish partial from high-efficacy degrader profiles	Linkers may be ranked highly despite limited degradation ceiling
Degradation kinetics	Time-dependent degradation measurements when available	Reflects how rapidly degradation occurs under cellular conditions	Supports activity-profile conditioning beyond static endpoints	Slow or transient degradation behavior may be missed
Assay metadata	Cell type, target abundance, incubation time, detection method	Preserves biological context behind degradation labels	Prevents inappropriate pooling of heterogeneous activity data	The model may learn confounded labels that do not generalize
Synthetic feasibility	Retrosynthetic accessibility, known linker motifs, reaction compatibility	Ensures generated hypotheses can be tested experimentally	Filters candidates before medicinal chemistry review	Outputs may be imaginative but impractical for synthesis
Model uncertainty	Applicability-domain score or confidence estimate	Communicates reliability for rare targets, uncommon E3 ligases, or sparse chemotypes	Supports cautious prioritization and human review	Unfamiliar POI–E3 pairs may be presented with false confidence

Generative Model Architecture

Conditional Generative Framework

A conditional generative framework would generate linkers while conditioning on the two warheads, the POI–E3 pair, desired ternary geometry, and degradation-related labels. A variational autoencoder could learn a latent space of linker chemistry, while a diffusion model could sample linker structures under spatial constraints inspired by equivariant 3D linker design [16]. PROTAC-specific reinforcement learning provides an alternative mechanism for optimizing generated linkers against structural and activity-guided objectives [7]. In this framework, degradation labels would guide the model conceptually, but generated molecules would remain hypotheses requiring computational and experimental evaluation.

Figure 1 illustrates the proposed geometry- and degradation-guided generative architecture for converting PROTAC warheads, ternary complex constraints, and degradation annotations into experimentally prioritized linker hypotheses.

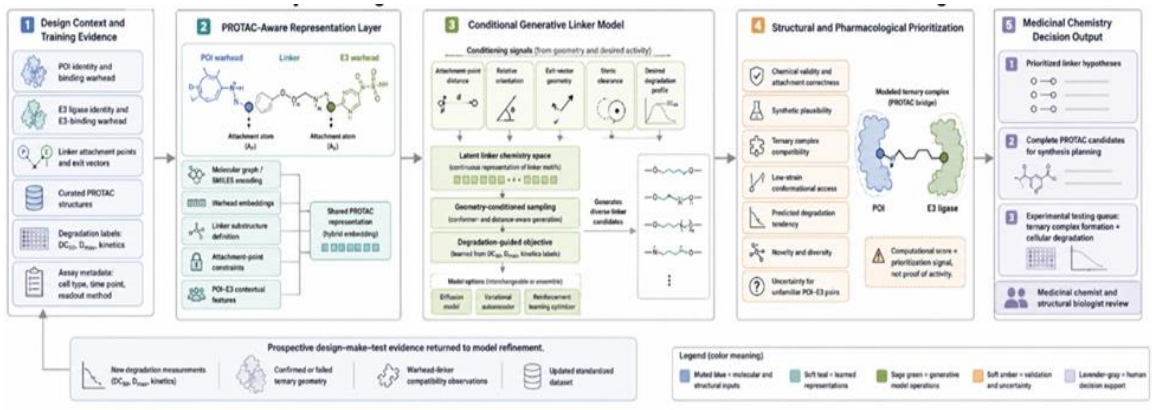

Figure 1. Geometry- and degradation-guided generative architecture for PROTAC linker design

Linker Generation and Warhead Attachment

The model would output a linker as a molecular graph, SMILES string, or 3D fragment with defined connection atoms. Link-INVENT and related linker-generation models provide a precedent for constructing linkers between molecular fragments under optimization constraints [8], while deep 3D linker design shows how spatial fragment relationships can shape generated connectivity [5]. For PROTACs, the virtual attachment step must preserve the chemistry of both warheads and maintain compatible exit vectors, because inappropriate connection geometry could disrupt binary binding or ternary complex formation. The final generated PROTAC would therefore be evaluated as a complete bifunctional molecule rather than as an isolated linker.

Training and Optimization

Training would aim to reconstruct or generate observed linkers from curated PROTAC examples while learning how linker chemistry relates to structural and degradation context. Auxiliary objectives could encourage generated linkers to remain compatible with modeled ternary complex geometries, using structural modeling frameworks such as PRosettaC or integrative ternary complex modeling as evaluators [12, 14]. Reinforcement learning could further optimize candidates toward desired properties, as demonstrated in linker-focused molecular design systems [8, 19]. Because prospective activity cannot be assumed from model output alone, optimization should be framed as prioritization of plausible candidates rather than prediction of confirmed degraders.

Incorporating Ternary Complex Geometry and Degradation Constraints

Geometry-Aware Conditioning

Geometry-aware conditioning would use the relative placement of the POI-binding and E3-binding warheads in a ternary complex model to guide linker generation. The distance, orientation, and exit-vector relationship between attachment points can be estimated from Rosetta-based or integrative ternary complex models [12, 14], while pose-scoring workflows can identify assemblies that are more consistent with productive PROTAC binding [15]. A generated linker should span the required geometry without imposing excessive strain or forcing either warhead away from its binding pose. In this setting, the model would not simply generate a linker of a chosen length, but would learn a geometry-conditioned distribution of linkers expected to support the desired ternary arrangement.

Table 2 shows the key structural and computational components used to guide geometry-aware conditioning for linker generation in PROTAC ternary complex modeling, integrating spatial constraints from warhead positioning, exit-vector alignment, and pose-scoring evaluations to ensure productive complex formation without introducing structural strain.

Table 2. Geometry-aware conditioning strategy for PROTAC linker design in ternary complex modeling

Component / Concept	Description	Role in Linker Generation	Typical Computational Approach
POI-binding warhead	Ligand that binds the protein of interest (POI)	Defines one anchor point of the linker	Docking / crystallographic pose refinement
E3 ligase-binding warhead	Ligand that binds the E3 ubiquitin ligase	Defines second anchor point of the linker	Docking / known ligand-bound structures
Attachment (exit) vectors	Points where linker connects to each warhead	Constrain feasible linker directionality	Vector extraction from bound poses
Inter-warhead distance	Spatial separation between anchor points in ternary complex	Determines approximate linker length requirement	Rosetta-based modeling / integrative modeling
Relative orientation	Angular relationship between warheads in complex	Influences linker flexibility and strain	Structural alignment / pose sampling

Ternary complex stability score	Energetic and geometric compatibility of assembled complex	Filters productive binding conformations	Pose scoring functions (e.g., Rosetta energy)
Linker geometry distribution	Set of feasible linker conformations conditioned on geometry	Guides generative model output space	Geometry-conditioned generative models

Degradation-Guided Generation

Degradation-guided generation would steer linker proposals using experimentally derived activity labels and surrogate predictors of PROTAC-induced degradation. DeepPROTACs introduced a deep learning framework for degradation prediction from PROTAC-related features [17], while ET-PROTACs and structure-informed ternary attention models illustrate how interaction-aware representations can support degradation modeling [20, 21]. Additional machine learning frameworks for PROTAC activity prediction suggest that degradation labels can be incorporated as ranking or guidance signals during design rather than treated as definitive experimental outcomes [22, 23]. A generative model could therefore be biased toward linkers associated with favorable degradation profiles while still requiring downstream validation.

Conformational Compatibility Check

Generated PROTACs should be subjected to conformational checks to assess whether the proposed linker can access a ternary-complex-compatible pose. Molecular dynamics and conformational studies indicate that linker plasticity can influence both solution conformations and ternary complex dissociation, while linker-dependent folding can affect cellular permeability and the exposure of polar functionality. Energy-landscape analyses of PROTAC-mediated protein-protein interactions further support the idea that generated designs should be evaluated across conformational states rather than by a single static pose. This check would filter out linkers that appear geometrically plausible in two dimensions but are unlikely to adopt a low-strain conformation compatible with productive complex formation.

Evaluating Generated Linkers

In-Silico Metrics

In-silico evaluation should assess chemical validity, novelty relative to known PROTAC linkers, structural diversity, attachment-point correctness, and predicted ternary complex compatibility. General linker-generation studies provide metrics for fragment connection and chemical plausibility [5], while PROTAC-specific generative and domain-adapted diffusion approaches illustrate how linker outputs can be evaluated in a chemical space relevant to bifunctional degraders [7, 24]. Structure-based filters from PRosettaC, Rosetta modeling, and ternary pose scoring can be used to assess whether generated molecules support plausible POI–PROTAC–E3 assemblies [12, 13, 15]. These metrics should be interpreted as prioritization tools rather than as evidence that a generated linker will produce cellular degradation.

Prospective Experimental Feedback

Prospective experimental feedback would involve selecting a small, diverse set of generated candidates for synthesis and biological testing, while avoiding claims that the model has already produced active degraders. Assays measuring ternary complex formation and cellular degradation can provide feedback labels that refine the generator, and degradation-prediction models such as DegradeMaster or SE(3)-PROTACs suggest how new measurements could be integrated into improved surrogate scoring [25]. Experimental results should be returned to the design loop as new evidence about linker geometry, warhead compatibility, and degradation behavior. In this closed-loop framework, the generative model would function as a hypothesis engine embedded within a design-make-test cycle.

Integration Into PROTAC Drug Discovery Workflow

Hit Expansion and Lead Optimization

For hit expansion, the model could generate focused linker libraries around a validated PROTAC by varying length, rigidity, polarity, and attachment geometry while preserving the two warheads. Linker-focused medicinal chemistry reviews emphasize that such changes can strongly affect degradation activity, selectivity, and physicochemical behavior [3, 4]. Generative linker methods can explore alternatives beyond familiar PEG and alkyl motifs, while PROTAC-specific generators can incorporate ternary complex and degradation constraints during prioritization [7, 8]. In lead optimization, this approach would help medicinal chemists compare structurally diverse linker hypotheses before committing to synthesis.

Target-Agnostic and Pan-E3 Design

A target-agnostic framework would train across multiple POIs and E3 ligases so that linker generation can adapt to different ternary complex geometries. Public PROTAC resources now include a broad range of targets, ligases, and compounds that can support cross-system learning [9-11], although coverage remains uneven across E3 families and target classes. Computational predictors such as DeepPROTACs, ET-PROTACs, and SE(3)-PROTACs suggest that molecular, structural, and interaction-aware representations can generalize degradation modeling across PROTAC contexts [17, 20]. A pan-E3 design strategy would therefore treat the E3 ligase as a conditioning variable, allowing the model to propose which E3-linker-POI combinations should be prioritized for a given target.

Evaluation Strategy

Retrospective Validation

Retrospective validation would test whether the model can recover plausible linker chemotypes for known PROTAC systems withheld from training. Generated linkers could be compared with reference linkers using chemical similarity, attachment geometry, and predicted ternary complex compatibility, following evaluation principles from deep linker design and reinforcement learning-based linker optimization [5, 8, 19]. PROTAC-specific benchmarks should also assess whether candidates preserve the identity and connectivity of both warheads, a challenge addressed by substructure-aware tools such as PROTAC-Splitter [18]. This evaluation would not prove prospective activity, but it would indicate whether the generator has learned recognizable design constraints from known degraders.

Prospective Synthesis and Testing

Prospective synthesis and testing would select a chemically diverse subset of generated PROTACs for experimental evaluation of ternary complex formation and cellular degradation. Structural studies of cooperative recognition and ligand-induced plasticity show why both complex formation and degradation should be assessed, because binary binding alone may not predict degradation outcome [1, 2]. Activity datasets and databases can then capture the resulting degradation annotations in a standardized form, improving future training data for generative and predictive models [9, 11]. The prospective study should be framed as validation of model-prioritized hypotheses, not as confirmation of performance numbers before experiments are conducted.

Model Robustness and Generalizability

Robustness should be evaluated by applying the model to target–E3 combinations that are chemically or structurally distinct from those most represented in training data. Cross-modal and structure-informed degradation predictors suggest that incorporating molecular and interaction-level information may improve generalization beyond simple chemical similarity [20, 21]. However, PROTAC activity prediction remains sensitive to assay context and data heterogeneity, as highlighted by multiple machine learning studies of PROTAC degradation and activity [22, 23]. A robust generative model should therefore expose uncertainty when proposing linkers for unfamiliar POI–E3 pairs rather than presenting such designs as equally reliable across all targets.

Table 3 provides a decision-governance framework for evaluating generated PROTAC linkers as prioritized experimental hypotheses rather than as computationally validated degraders.

Table 3. Evaluation and Decision-Governance Framework for Generated PROTAC Linker Hypotheses

Evaluation layer	Core question	Recommended assessment	Decision value for medicinal chemistry	Interpretation boundary
Chemical validity	Is the generated linker chemically coherent and valence-correct?	Structure sanitization, atom/bond validity, attachment-point correctness	Removes impossible structures before downstream scoring	Valid chemistry does not imply biological activity
Warhead attachment compatibility	Can the linker be attached without disrupting either warhead?	Exit-vector inspection, local steric analysis, attachment-site preservation	Protects POI and E3 binding assumptions	Computational attachment checks may miss altered binding affinity
Synthetic plausibility	Can the linker-containing PROTAC be realistically synthesized?	Retrosynthetic route availability, reaction compatibility, building-block availability	Prioritizes candidates that can enter design–make–test cycles	Synthetic accessibility scores are not substitutes for expert route planning
Geometry compatibility	Does the linker span the required ternary complex arrangement?	Attachment-point distance, relative orientation, steric clearance, pose-scoring	Selects candidates consistent with modeled POI–E3 geometry	Modeled geometry may differ from cellular ternary complex states
Conformational accessibility	Can the PROTAC adopt a low-strain ternary-compatible conformation?	Conformer ensemble analysis, strain estimation, molecular dynamics when appropriate	Filters designs that only work in unrealistic static poses	Conformational sampling remains approximate for large flexible PROTACs
Ternary complex plausibility	Does the full PROTAC support productive POI–E3 proximity?	PRosettaC-like or Rosetta-based modeling, ternary pose scoring, interface analysis	Prioritizes linkers likely to support cooperative recognition	Plausible complex formation does not guarantee ubiquitination
Degradation prediction	Is the candidate associated with favorable degradation behavior?	Surrogate degradation predictors, DC50/Dmax-conditioned ranking, activity-profile scoring	Helps rank candidates by intended pharmacological effect	Predicted degradation must not be reported as confirmed degradation
Novelty and diversity	Does the model explore linker space beyond empirical motifs?	Similarity to known PROTAC linkers, scaffold diversity, property distribution	Expands medicinal chemistry options beyond PEG/alkyl enumeration	Novelty without feasibility or geometry is not useful

Physicochemical risk	Could the linker worsen permeability, polarity, or molecular size concerns?	cLogP, polar surface area, rotatable bonds, conformational shielding indicators	Flags developability risks before synthesis	Simple property filters may be unreliable for folded or shielded PROTAC conformations
Applicability-domain uncertainty	Is the model operating within familiar POI–E3 and linker chemistry space?	Uncertainty estimates, nearest-neighbor similarity, target/E3 coverage analysis	Prevents overconfidence for rare targets or underrepresented E3 ligases	High uncertainty should trigger cautious review, not automatic rejection
Experimental prioritization	Which candidates should move into synthesis and testing?	Multi-criteria ranking across geometry, degradation prediction, diversity, feasibility, and uncertainty	Converts generative outputs into a practical testing queue	Ranking supports selection; it does not validate efficacy
Prospective feedback	How should experimental results improve the next model cycle?	Capture ternary complex data, degradation assays, failed designs, assay metadata, and synthesis outcomes	Turns experimental results into new training and refinement evidence	Feedback quality depends on standardized reporting and negative-result retention

Limitations

Dependency on Warhead Binding Affinity

The proposed model assumes that the two warheads maintain productive binding after linker attachment, but linker chemistry can perturb binary affinity, exit vectors, and local binding-site interactions. Structural studies show that degradation can depend on induced ternary contacts and ligand positioning rather than on isolated ligand affinity alone [1, 2]. Computational ternary complex modeling can identify some incompatible poses, but it may not fully capture changes in warhead binding caused by new attachment chemistry [12, 15]. Therefore, warhead-linker compatibility remains a critical limitation that should be addressed through explicit binding-pose evaluation and experimental follow-up.

Limited Training Data and Generalization

PROTAC data remain limited and heterogeneous relative to conventional small-molecule datasets, especially for uncommon E3 ligases, underexplored targets, and standardized degradation kinetics. Although PROTAC-DB and its updates provide essential curated resources [9-11], model training may still be biased toward well-studied CRBN and VHL chemotypes. Degradation predictors such as DeepPROTACs, DegradeMaster, and SE(3)-PROTACs illustrate the promise of machine learning, but they also depend on the breadth and consistency of available activity annotations [17, 25]. Consequently, a generative linker model should be presented as a decision-support tool whose reliability depends on continued expansion and standardization of PROTAC datasets.

Conclusion

A generative model for PROTAC linker design can be framed as a structure-informed system that learns how linker chemistry relates to ternary complex geometry and degradation intent. Rather than generating arbitrary bifunctional molecules, the model would condition on the POI, E3 ligase, warheads, attachment points, and desired degradation behavior. This conceptual design places linker generation at the intersection of molecular representation learning, structural modeling, and targeted protein degradation biology. It offers a principled way to move from empirical linker variation toward computationally prioritized linker hypotheses.

The strength of this framework lies in its ability to combine geometric constraints with degradation-aware guidance. Ternary complex information can help the model avoid linkers that are chemically valid but structurally implausible, while degradation labels can bias generation toward designs consistent with the intended pharmacological effect. By exploring linker chemical space more broadly than conventional medicinal chemistry enumeration, the model could support creative yet constrained design. Such a workflow would help prioritize molecules that deserve synthesis and testing.

Important challenges remain before this approach can become a routine component of PROTAC discovery. Training data are sparse and heterogeneous, warhead-linker attachment can alter binding in ways that are difficult to predict, and structural models may not capture all conformational states relevant to cellular degradation. Prospective experimental validation is therefore essential, especially for new POI–E3 combinations and less common E3 ligases. The model should be evaluated as a hypothesis-generating system rather than as a substitute for biochemical and cellular assays.

Future progress will depend on open-source models, transparent benchmarking, and larger standardized datasets that report both chemical structures and degradation assay context. Community resources should capture ternary complex structures, modeled geometries, linker annotations, and harmonized degradation endpoints. As these data improve, generative models could become increasingly useful for designing PROTAC linkers that are structurally plausible, synthetically accessible, and biologically purposeful. This would help establish a new generation of computational tools for targeted protein degradation.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Gadd MS, Testa A, Lucas X, Chan KH, Chen W, Lamont DJ, et al. Structural basis of PROTAC cooperative recognition for selective protein degradation. Nat Chem Biol. 2017;13(5):514-21.
2. Nowak RP, DeAngelo SL, Buckley D, He Z, Donovan KA, An J, et al. Plasticity in binding confers selectivity in ligand-induced protein degradation. Nat Chem Biol. 2018;14(7):706-14.
3. Troup RI, Fallan C, Baud MG. Current strategies for the design of PROTAC linkers: a critical review. Explor Target Antitumor Ther. 2020;1(5):273-312.
4. Bemis TA, La Clair JJ, Burkart MD. Unraveling the role of linker design in proteolysis targeting chimeras: Miniperspective. J Med Chem. 2021;64(12):8042-52.
5. Imrie F, Bradley AR, van der Schaar M, Deane CM. Deep generative models for 3D linker design. J Chem Inf Model. 2020;60(4):1983-95.
6. Imrie F, Hadfield TE, Bradley AR, Deane CM. Deep generative design with 3D pharmacophoric constraints. Chem Sci. 2021;12(43):14577-89.
7. Li B, Ran T, Chen H. 3D based generative PROTAC linker design with reinforcement learning. Brief Bioinform. 2023;24(5):bbad323.
8. Guo J, Knuth F, Margreitter C, Janet JP, Papadopoulos K, Engkvist O, et al. Link-INVENT: generative linker design with reinforcement learning. Digital Discovery. 2023;2(2):392-408.
9. Weng G, Shen C, Cao D, Gao J, Dong X, He Q, et al. PROTAC-DB: an online database of PROTACs. Nucleic Acids Res. 2021;49(D1):D1381-7.
10. Weng G, Cai X, Cao D, Du H, Shen C, Deng Y, et al. PROTAC-DB 2.0: an updated database of PROTACs. Nucleic Acids Res. 2023;51(D1):D1367-72.
11. Ge J, Li S, Weng G, Wang H, Fang M, Sun H, et al. PROTAC-DB 3.0: an updated database of PROTACs with extended pharmacokinetic parameters. Nucleic Acids Res. 2025;53(D1):D1510-5.
12. Zaidman D, Prilusky J, London N. PRosettaC: Rosetta based modeling of PROTAC mediated ternary complexes. J Chem Inf Model. 2020;60(10):4894-903.
13. Bai N, Miller SA, Andrianov GV, Yates M, Kirubakaran P, Karanicolas J. Rationalizing PROTAC-mediated ternary complex formation using Rosetta. J Chem Inf Model. 2021;61(3):1368-82.
14. Weng G, Li D, Kang Y, Hou T. Integrative modeling of PROTAC-mediated ternary complexes. J Med Chem. 2021;64(21):16271-81.
15. Liao J, Nie X, Unarta IC, Ericksen SS, Tang W. In silico modeling and scoring of PROTAC-mediated ternary complex poses. J Med Chem. 2022;65(8):6116-32.
16. Igashov I, Stärk H, Vignac C, Schneuing A, Satorras VG, Frossard P, et al. Equivariant 3D-conditional diffusion model for molecular linker design. Nat Mach Intell. 2024;6(4):417-27.
17. Li F, Hu Q, Zhang X, Sun R, Liu Z, Wu S, et al. DeepPROTACs is a deep learning-based targeted degradation predictor for PROTACs. Nat Commun. 2022;13(1):7133.
18. Ribes S, Zhang R, Cropsal T, Källberg A, Tyrchan C, Nittinger E, et al. PROTAC-Splitter: a machine learning framework for automated identification of PROTAC substructures. J Cheminform. 2026;18(1):30.
19. Tan Y, Dai L, Huang W, Guo Y, Zheng S, Lei J, et al. DRlinker: deep reinforcement learning for optimization in fragment linking design. J Chem Inf Model. 2022;62(23):5907-17.
20. Cai L, Yue G, Chen Y, Wang L, Yao X, Zou Q, et al. ET-PROTACs: modeling ternary complex interactions using cross-modal learning and ternary attention for accurate PROTAC-induced degradation prediction. Brief Bioinform. 2025;26(1):bbae654.
21. Chen Z, Gu C, Tan S, Wang X, Li Y, He M, et al. Interpretable PROTAC degradation prediction with structure-informed deep ternary attention framework. Adv Sci. 2025;12(47):e08138.
22. Ribes S, Nittinger E, Tyrchan C, Mercado R. Modeling PROTAC degradation activity with machine learning. Artif Intell Life Sci. 2024;6:100104.
23. Abouzied AS, Alshammari B, Kari H, Huwaimel B, Alqarni S, Kassab SE. AI-DPAPT: a machine learning framework for predicting PROTAC activity. Mol Divers. 2025;29(4):2995-3007.
24. Song Z, Meng Z, Hernández-Lobato JM. Domain-adapted diffusion model for PROTAC linker design through the lens of density ratio in chemical space. In: Forty-second International Conference on Machine Learning (ICML); 2025.
25. Liu J, Roy MJ, Isbel L, Li F. Accurate PROTAC-targeted degradation prediction with DegradeMaster. Bioinformatics. 2025;41(Suppl 1):i342-51.