Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of structurally diverse Tau-Targeting PROTACs using Integrated Computational Prioritization framework

Banesh Sooram
Karolinska Institutet
Bengt Winblad
Karolinska Institutet
Lars O Tjernberg
Karolinska Institutet
Lisha Wang
lisha.wang@ki.se

Karolinska Institutet

Article

Keywords: Tau, proteolysis-targeting chimera, targeted protein degradation, BT-MPO, ternary-complex modeling, central nervous system drug discovery

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11427623/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Identification of structurally diverse Tau-Targeting PROTACs using Integrated Computational Prioritization framework

Banesh Sooram[1#], Bengt Winblad[1,2], Lars Tjernberg[1], Lisha Wang[1*]

[1]Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Solna, Sweden

[2]Theme Inflammation and Aging, Karolinska University Hospital, Huddinge, Sweden

***Corresponding author**: Lisha Wang, Lisha.wang@ki.se

BioClinicum J9:20, Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, 171 64 Solna, Sweden.

#Co-corresponding author: Banesh Sooram, banesh.sooram@ki.se

BioClinicum J9:20, Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, 171 64 Solna, Sweden.

Abstract:

Tau-targeting proteolysis-targeting chimeras (PROTACs) offer a strategy for eliminating intracellular pathological Tau, but their development requires simultaneous optimization of central nervous system properties and ternary-complex formation. We developed an integrated computational framework for prioritizing Tau-targeting PROTACs from a 16,400-member combinatorial library comprising diverse Tau-binding ligands, linkers, cereblon (CRBN) recruiters, and exit vector variants. Candidates were evaluated using predicted stability and pharmacokinetic properties and ranked using a Brain-Targeted Multi-Parameter Optimization (BT-MPO) score. Benchmarking and weight-sensitivity analyses supported the robustness of compound ranking, while extended-connectivity fingerprint-based chemical-space analysis guided the selection of high-ranking candidates with broad structural representation. Ternary-complex modeling demonstrated that linker composition and conformational constraints influenced modeled Tau-PROTAC-CRBN assembly beyond linker length alone. Five leading candidates were subsequently examined using molecular-dynamics simulations, protein-ligand interaction analysis, relative MM/PBSA energetics, free-energy landscapes, and dynamic cross-correlation analysis. We identified a PROTAC candidate with the most favorable pharmacokinetic, and modeled ternary-complex stability profile. In summary, this study provides a systematic strategy for combining central nervous system (CNS)-oriented multiparameter prioritization with structural and dynamic analyses to identify diverse Tau-targeting PROTAC candidates for subsequent experimental evaluation.

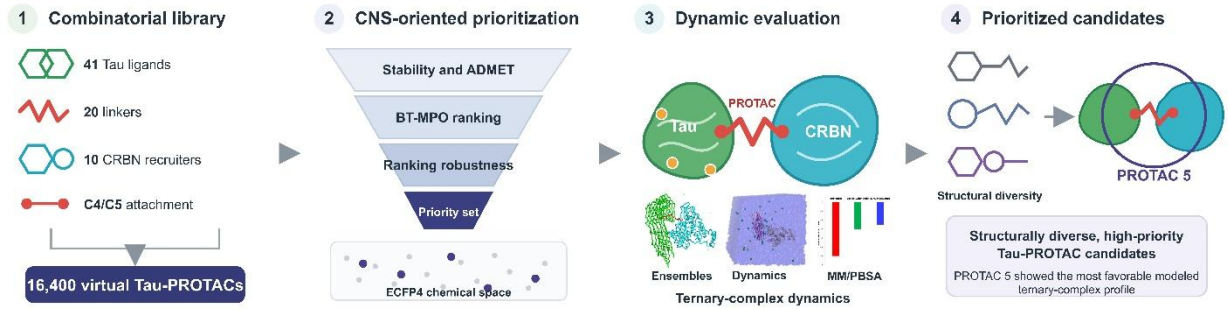
Scientific contribution:

This study introduces an integrated cheminformatics framework for prioritizing CNS-oriented PROTACs from large combinatorial libraries. The framework combines transparent multiparameter scoring, external benchmarking, weight-sensitivity analysis, ECFP4-based chemical-space mapping, and ensemble-based ternary-complex evaluation. Applied to a 16,400-member Tau-targeting PROTAC library, the approach produced a structurally diverse set of high-priority candidates while preserving broad chemical-space representation. The accompanying

scoring definitions, code, and molecular datasets provide a reproducible foundation for adapting the workflow to other targeted protein-degradation systems.

Keywords: Tau; proteolysis-targeting chimera; targeted protein degradation; BT-MPO; ternary-complex modeling; central nervous system drug discovery

Graphical Abstract:



Integrated cheminformatics preserves structural diversity and identifies testable Tau–PROTAC candidates.

1. Introduction:

Proteolysis-targeting chimeras (PROTACs) are heterobifunctional molecules that recruit a protein of interest to an E3 ubiquitin ligase, thereby promoting target ubiquitination and proteasomal degradation [1] . A PROTAC comprises a target-binding ligand, an E3-ligase-recruiting ligand, and a chemical linker connecting the two binders. Formation of a productive protein of interest-PROTAC-E3 ligase ternary complex is an essential step in this process because degradation depends not only on binary ligand affinity but also on the orientation of the recruited proteins, and conformational ensemble of the assembled complex. These properties influence the presentation of accessible target lysine's for ubiquitination. PROTACs provide an event-driven pharmacological modality capable of eliminating the complete target protein, including nonenzymatic and scaffolding functions that may not be addressed by conventional occupancy-based inhibitors [2] .

PROTACs also frequently occupy beyond-Rule-of-Five (bRo5) chemical space and exhibit molecular weights, polarities, hydrogen-bonding capacities, and conformational flexibility beyond the typical small molecule drugs. These properties create substantial challenges for aqueous solubility, membrane permeability, oral exposure, metabolism, and intracellular accumulation [3]. The difficulty is amplified for central nervous system applications, where compounds must additionally achieve sufficient brain exposure across the blood-brain barrier (BBB) while avoiding excessive efflux and peripheral distribution. Brain delivery of PROTACs is therefore limited by the size, polarity, and flexibility that enable simultaneous engagement of the target protein and E3 ligase [4] . CNS-directed degrader design requires balancing multiple, often competing objectives. Multiparameter optimization provides a useful conceptual basis for such prioritization by integrating several physicochemical properties through continuous desirability functions instead of relying exclusively on individual hard cutoffs. The established Central Nervous System Multiparameter Optimization (CNS MPO) framework similarly showed that holistic alignment of lipophilicity, molecular weight, polarity, hydrogen-bonding capacity, and ionization-related properties can be more informative than single-parameter rules [5] .

Alzheimer's disease is characterized by amyloid-β plaques and neurofibrillary tangles containing pathologically modified Tau. Tau hyperphosphorylation, mislocalization, and aggregation contribute to neurodegeneration, while tangle burden correlates with dementia severity, supporting Tau as a disease-modifying target [6] . Targeted protein degradation offers a strategy for removing pathogenic intracellular Tau rather than inhibiting an individual interaction or modification. Its feasibility has been demonstrated using multiple degrader architectures, including TH006, a Keap1–Cul3-recruiting peptide PROTAC, the CRBN-recruiting degrader QC-01-175, and the VHL-recruiting compounds C004019 and C8. These molecules reduced Tau in cellular, patient-derived neuronal, or animal models, with selected compounds also improving synaptic or cognitive outcomes in vivo ([7], [8], [9], [10], [11], [12]. Collectively, these studies establish the feasibility of Tau-targeted degradation and motivate computational exploration of broader PROTAC chemical space before resource-intensive synthesis and experimental evaluation.

Computational approaches can reduce the experimental design space by evaluating linker conformations, protein-protein orientations, ternary-complex ensembles, and molecular dynamics before synthesis. Drummond and Williams introduced computational methods for generating PROTAC-mediated ternary-complex ensembles, which were later extended by Drummond et al. through the specialized clustering procedure implemented in Method 4B [13], [14]. Molecular-dynamics (MD) studies have further emphasized that ternary-complex behavior depends on linker length and conformation, interfacial residues, E3-ligase identity, and coordinated motions that cannot be captured fully by a single static pose [15]. Recent computational reviews describe a rapidly expanding toolkit that includes cheminformatics, ternary-complex modeling, molecular dynamics, machine learning, activity prediction, and ADMET assessment [16]. Nevertheless, these approaches have largely been developed and benchmarked using oncology-associated or general degrader systems. Their application to Tau requires additional consideration of CNS-related physicochemical properties, BBB exposure potential, and ternary-complex geometries that position solvent-exposed Tau lysines near the recruited E3-ligase machinery.

Existing Tau-degrader studies have generally examined focused compound series rather than systematically enumerated libraries spanning diverse Tau ligands, linkers, CRBN recruiters, and attachment geometries. Because productive degradation cannot be reliably inferred from individual physicochemical descriptors or static structural models alone, candidate selection requires integrated assessment of CNS-oriented properties and ternary-complex behavior. Accordingly, in the present study, we developed an integrated computational framework for the design and relative prioritization of Tau-targeting PROTACs intended for CNS application. A 16,400-member virtual library was generated by systematically combining 41 Tau-binding ligands, 20 linker scaffolds, 10 CRBN-recruiting ligand scaffolds, and two CRBN-ligand exit vectors. The library was evaluated using plasma- and chemical-stability prediction, complementary pharmacokinetic assessment, stringent CNS-oriented reference criteria, a Brain-Targeted Multi-Parameter Optimization ranking framework, and Extended-Connectivity Fingerprint with a diameter of 4 (ECFP4)-based chemical-space analysis. Prioritized candidates were subsequently examined using MOE Method 4B ternary-complex modeling, proximal-lysine assessment, two independent 100-ns molecular-dynamics simulations, MM/PBSA energetic analysis, principal-component-based free-energy landscapes, and dynamic cross-correlation matrices. The objective was not to assign experimental brain penetration or Tau-degradation activity computationally, but to identify structurally diverse, best-balanced candidates and linker-dependent ternary-complex hypotheses suitable for subsequent synthesis and experimental validation.

2. Methods:

Collection of E3 ligase and target protein binding ligands: Tau-binding ligands were collected from the BindingDB database (https://www.bindingdb.org/rwd/bind/index.jsp) following a search for compounds reported to interact with human Tau (MAPT). Compounds were manually curated to remove duplicates and closely related analogues, and representative ligands with defined chemical structures and reported experimental binding data, including KiK_iKi values, were retained to maximize scaffold diversity. CRBN E3-ligase recruiting ligands were selected from clinically advanced and extensively validated CRBN-based PROTACs reported in the literature.

Emphasis was placed on recruiting moieties derived from thalidomide, lenalidomide, pomalidomide, and related CRBN-binding analogs. To ensure adequate exploration of chemical space, ligands were selected to represent diverse attachment vectors, and linker structures with C4 and C5 exit positions. The final ligand set was curated to maximize structural diversity across both Tau-targeting warheads and CRBN-recruiting ligands, and linkers thereby enabling the generation of a chemically diverse PROTAC library for subsequent combinatorial design and computational screening.

PROTAC Library Design: A virtual PROTAC library was generated using a custom RDKit-based Python workflow. Tau-binding ligands (n = 41), CRBN-recruiting ligands (n = 10) and linker scaffolds (n = 20) were selected to maximize structural diversity and exit vector variability. To enable automated assembly, predefined conjugation sites were designated using dummy atoms within the SMILES representations. Attachment points on Tau-binding ligands were denoted as [*:1], whereas attachment points on CRBN ligands were denoted as [*:2]. Linker structures were encoded with complementary terminal attachment points ([*:1] and [*:2]) to facilitate systematic connection between the Tau warhead and the E3-ligase recruiting moiety. For the CRBN ligands, two alternative linker attachment positions corresponding to the C4 and C5 substitution sites were considered. Inclusion of multiple exit vectors was intended to capture positional effects on PROTAC geometry, ternary-complex formation, and degradation potential. The total library size was calculated according to:

$$N_{PROTAC} = N_{Tau} \times N_{Linker} \times N_{CRBN}$$

This resulted in 8,200 unique PROTAC structures for each CRBN exit vector position (C4 or C5). All enumerated compounds were standardized and exported in canonical SMILES format for subsequent analysis.

Stability Prediction: Plasma and chemical stability were predicted from canonical SMILES using PredPS and DeepChemStable, respectively. PredPS, an attention-based graph neural network, classifies susceptibility to plasma-mediated degradation and reported an AUC of 0.901 and accuracy of 83.5% under five-fold cross-validation [17] . DeepChemStable, an attention-based graph convolutional model trained on experimental decomposition data, predicts stability under storage conditions and reported an AUC of 0.847 and accuracy of 79.1% under 10-fold stratified cross-validation [18] . PROTACs predicted to be stable by both models were retained for subsequent analyses.

Pharmacokinetic and Blood–Brain Barrier Permeability Prediction

PROTAC candidates retained after stability screening were evaluated using Deep-PK [19] , SwissADME [20] , and ADMETlab 3.0 [21] . These complementary platforms provided physicochemical, pharmacokinetic, toxicity-related, and CNS-exposure predictions for subsequent candidate assessment. Because individual endpoints were not uniformly available or directly comparable across all platforms, the outputs were considered collectively as complementary evidence rather than combined as interchangeable numerical estimates. The resulting predictions were subsequently evaluated using the CNS-oriented reference criteria and BT-MPO framework described below.

CNS-Focused Filtering and Brain-Targeted Multi-Parameter Optimization Framework (BT-MPO)

Candidates retained after stability screening underwent criterion-level profiling and continuous multiparameter ranking. Criterion-level profiling evaluated eight CNS-oriented properties: cLogP 3.0-4.0 as a lipophilicity proxy, TPSA ≤150 Å², MW ≤800 Da, HBA ≤12, HBD ≤2, ESOL-derived aqueous solubility ≥0.001 mg/mL, RDKit synthetic-accessibility score <5, and satisfaction of the heuristic BBB-suitability criterion. ESOL logS values were converted from mol/L to mg/mL using molecular weight. These criteria were used to quantify overall coverage and identify prevalent predicted liabilities.

The BT-MPO score integrated a modified CNS MPO-like score, heuristic BBB-suitability index, heuristic oral-developability index, solubility desirability score, quantitative estimate of drug-likeness, MW compliance, and a structural-alert and safety-proxy penalty. Continuous descriptors were transformed to values between 0 and 1 using piecewise-linear desirability functions.

Descriptor selection was informed by established CNS MPO, BBB Score, QED, and heterobifunctional degrader developability frameworks [22], [23], [24], [25] . The precise desirability ranges and coefficients were defined as a priori as an expert-guided, PROTAC-adapted ranking scheme and were not fitted to experimental data. Accordingly, BT-MPO was used for relative candidate prioritization rather than as a validated predictor of BBB penetration, oral bioavailability, toxicity, or Tau degradation. Complete desirability functions, thresholds, structural alerts, and proxy definitions are provided in the Supporting Information.

Modified CNS MPO-Like Score

The modified CNS MPO-like score incorporated equally weightsfor MW, cLogP, TPSA, HBD, HBA, and RotB:

$$\boldsymbol{CNS_{\mathrm{MPO-like}} = 6 \times \mathrm{mean}\left(D_{\mathrm{MW}}, D_{\mathrm{clogP}}, D_{\mathrm{TPSA}}, D_{\mathrm{HBD}}, D_{\mathrm{HBA}}, D_{\mathrm{RotB}}\right)} \quad \textbf{---(1)}$$

The score ranged from 0 to 6. It was designated “modified CNS MPO-like” because its descriptor set and PROTAC-adapted desirability ranges differed from the published CNS MPO model [22] .

Heuristic BBB-Suitability Score

A heuristic BBB-suitability index was calculated from from following equation.

$$\begin{aligned} \boldsymbol{BBB_{\mathrm{score}} =} \quad & \boldsymbol{0.18D_{\mathrm{MW}} + 0.22D_{\mathrm{TPSA}} + 0.18D_{\mathrm{cLogP}} + 0.10D_{\mathrm{HBD}}} \\ & \boldsymbol{+0.10D_{\mathrm{HBA}} + 0.10D_{\mathrm{RotB}} + 0.06D_{\mathrm{charge}} + 0.06D_{\mathrm{Fsp3}}} \end{aligned} \quad \textbf{---(2)}$$

The descriptor classes were selected based on their established qualitative relevance to CNS exposure, including molecular size, polarity, lipophilicity, hydrogen-bonding capacity, ionization, flexibility, and molecular shape [22], [23], [24]. The author-defined coefficients emphasized TPSA, MW, and cLogP and summed to 1.0. The index was used for relative ranking and was not interpreted as a validated predictor of brain penetration.

Heuristic Oral-Developability Score

The heuristic oral-developability index combined normalized desirabilities for MW, cLogP, TPSA, HBD, HBA, RotB, and ESOL-predicted solubility:

$$Oral_{\text{score}} = 0.18D_{\text{MW}} + 0.18D_{\text{cLogP}} + 0.18D_{\text{TPSA}} + 0.14D_{\text{HBD}} + 0.14D_{\text{HBA}} + 0.10D_{\text{RotB}} + 0.08D_{\text{solubility}} \quad \text{---(3)}$$

The coefficients were specified as a priori for relative ranking and were not calibrated against measured oral bioavailability. This index was therefore interpreted as a comparative developability measure rather than a quantitative prediction of oral exposure.

Structural-Alert and Safety-Proxy Assessment

Potential medicinal-chemistry and assay-interference liabilities were assessed using the PAINS, Brenk, and NIH filter catalogs implemented in RDKit, together with an author-defined panel of SMARTS-based structural alerts. PAINS filters identify substructures associated with frequent or promiscuous assay activity, while structural-alert methods support conservative identification of potentially reactive or otherwise problematic motifs [26], [27], [28], [29]. Alerts were treated as computational triage indicators rather than evidence of compound-specific toxicity.

A structural-alert and safety-proxy penalty was calculated as:

$$Toxicity = 0.15N_{\text{PAINS}} + 0.10N_{\text{Brenk}} + 0.20N_{\text{toxophore}} + 0.20I_{\text{hERG}} + 0.15I_{\text{DILI}} + 0.10I_{\text{Pgp}} + 0.10I_{\text{CYP}} \quad \text{---(4)}$$

The coefficients were defined a priori to assign greater influence to custom reactive-group alerts and the hERG-related proxy than to general medicinal-chemistry, transporter, or CYP-related flags. The final penalty was restricted to 0-1 and was not interpreted as a validated toxicity prediction.

BT-MPO Score

The BT-MPO score was calculated as:

$$BT-MPO\ score = 0.34(BBB_{\text{score}}) + 0.18\left(\frac{CNS_{\text{MPO-like}}}{6}\right) + 0.18(Oral_{\text{score}}) + 0.12(Solubility_{\text{score}}) + 0.08(QED) + 0.10(MW_{\leq 800}) - 0.25(Toxicity)$$

--- (5)

All positive components were normalized to 0–1 before combination, and the final score was restricted to the same interval. $MW_{\leq 800}$was coded as 1 for compounds with MW ≤800 Da and 0 otherwise. Higher BT-MPO values indicated closer alignment with the prespecified CNS-oriented and developability objectives.

The weights reflected the intended prioritization hierarchy of the study. Heuristic BBB suitability received the largest positive contribution, followed by the modified CNS MPO-like and oral-developability indices. Solubility, MW compliance, and QED provided complementary developability contributions, while the structural-alert and safety-proxy term was applied as an independent penalty. Because no compound satisfied all eight reference criteria, compounds were selected by relative BT-MPO ranking rather than complete threshold compliance. Candidates were ranked separately within the C4 and C5 libraries, and the 50 highest-ranked compounds from each library were retained. Dependence of candidate ranking on the author-defined coefficients was evaluated by equal-weight comparison and Monte Carlo weight-perturbation analysis, as described below.

Benchmarking and BT-MPO framework Weight-Sensitivity Analysis

The BT-MPO framework was evaluated using 362 Tau-targeting PROTACs extracted from PROTAC-DB 3.0 [30] and an external set of reported CNS-directed PROTACs. All compounds were processed using the same descriptors, desirability functions, and coefficients used for prospective prioritization. PROTAC-DB 3.0 contains published PROTAC structures and associated pharmacokinetic information.

Weight sensitivity was assessed using an equal-weight comparator and Monte Carlo perturbation. In the equal-weight model, each of the six positive components received a weight of $1/6$, while the original safety-penalty coefficient was retained. Agreement with the predefined model was evaluated using Spearman rank correlation and overlap among the top-20, top-50, and top-100 compounds; only top-20 overlap was evaluated for the smaller CNS-directed set.

For Monte Carlo analysis, positive weights were independently varied by ±10%, ±20%, or ±30% and renormalized to sum to 1.0, while the safety-penalty coefficient was perturbed independently. Five thousand alternative models were generated at each level. Rank correlation and retention of the original top-ranked compounds were summarized as medians and 2.5th–97.5th percentile intervals.

Properties associated with prioritization were examined by comparing the upper and lower BT-MPO quartiles of the Tau-targeting set using two-sided Mann-Whitney U tests. The same properties were compared between the Tau-targeting and CNS-directed sets to provide external chemical-space context. Because experimental brain exposure was not uniformly available, the CNS-directed compounds were not treated as validated BBB-positive controls.

ECFP4-Based Chemical-Space and Regional-Coverage Analysis

Chemical-space analysis assessed whether BT-MPO prioritization retained structural diversity within the post-stability library. ECFP4 fingerprints were selected because they encode local atom-centered substructures and are widely used for molecular similarity, clustering, and structure–activity analysis [31]. Valid canonical SMILES were converted to RDKit molecules and represented as 2,048-bit Morgan fingerprints with radius 2.

Fingerprints from the complete-case C4 and C5 populations were combined into a common dataset of 6,445 PROTACs. UMAP was performed using 25 nearest neighbors, a minimum distance of 0.15, Jaccard distance, two dimensions, and a random seed of 42. BT-MPO scores were mapped onto the resulting embedding, and the 50 highest-ranked candidates from each attachment variant were overlaid using distinct symbols.

A single Uniform Manifold Approximation and Projection model was fitted to the combined binary fingerprint matrix using 25 nearest neighbors, a minimum distance of 0.15, the Jaccard distance metric, two output dimensions, and a fixed random seed of 42. The resulting two-dimensional coordinates were designated UMAP1 and UMAP2. Combining both attachment-variant libraries before embedding ensured that C4 and C5 compounds were represented within the same coordinate system. The BT-MPO score was mapped to the common UMAP embedding as a continuous color variable. The final 50 C4 and 50 C5 candidates were overlaid on the complete chemical-space representation using distinct attachment-specific symbols. Regions within the common UMAP embedding were identified using HDBSCAN with a minimum sample number of 5. HDBSCAN was applied to the common UMAP embedding using min_samples = 5 and

min_cluster_size = 80, equivalent to approximately one-eightieth of the 6,445-compound population. Points labeled -1 were treated as noise and were not counted as a separate region.

Regional coverage was calculated as:

$$\textbf{Regional coverage}\ (\%) = \frac{N_{\text{regions containing at least one prioritized candidate}}}{N_{\text{total identified regions}}} \times 100 \qquad \text{---(6)}$$

A total of 15 HDBSCAN-defined UMAP regions were identified. Regional coverage was calculated separately for the C4 and C5 prioritized subsets and for the combined selection.

Analysis of PROTAC-induced ternary complex formation: BT-MPO-prioritized candidates were modeled as Tau-PROTAC-CRBN ternary complexes using Method 4B in the MOE Targeted Protein Degradation module [32] . Method 4B samples alternative PROTAC conformations and protein–protein orientations and clusters the resulting ternary-complex solutions. Complexes were ranked by Double Cluster Population (DCP), with higher values indicating more consistently populated modeled configurations. Top-ranked poses were assessed for protein–PROTAC and protein–protein contacts and for solvent-exposed Tau lysines proximal to the E3-facing region. Five candidates were selected based on DCP, ternary-complex geometry, interfacial contacts, and proximal-lysine coverage. The selected complexes were prepared using MOE QuickPrep and exported in PDB format for molecular-dynamics simulations and comparative MM/PBSA energetic analysis.

Molecular dynamics simulations: The highest-ranked ternary-complex pose exhibiting both a favorable DCP score and the presence of accessible Tau lysine residues near the predicted ubiquitination interface was selected as the starting structure for molecular dynamics (MD) simulations. Protein coordinates for Tau and CRBN were obtained from the Protein Data Bank (PDB IDs: 7NRV and 4CI3, respectively). GROMOS compatible topology parameters for the PROTAC molecules were generated using the Automated Topology Builder (ATB) (https://atb.uq.edu.au/) web server. The generated PROTAC topologies were integrated with the corresponding protein topologies to construct complete Tau-PROTAC-CRBN ternary-complex systems.

All-atom MD simulations were performed using GROMACS v2023.4 [33] . The GROMOS54A7 force field [34] was employed for proteins and ligands because of its compatibility with ATB-generated parameters and its extensive application in biomolecular simulations. The ternary complexes were solvated in a cubic simulation box with a minimum distance of 20 Å between the solute and the box boundaries using the TIP3P water model. Sodium and chloride ions were added to neutralize the systems and to achieve a physiological ionic strength of approximately 150 mM. Energy minimization was performed using the steepest-descent algorithm until the maximum force acting on any atom was below 1000 kJ mol^{-1} nm^{-1}. The minimized systems were subsequently equilibrated in two stages. First, a 100 ps NVT simulation was carried out at 310 K with positional restraints applied to heavy atoms using a force constant of 1000 kJ mol^{-1} nm^{-2}. This was followed by a 1 ns NPT equilibration at 310 K and 1 atm while maintaining positional restraints. Temperature was controlled using the velocity-rescale thermostat, and pressure was maintained using the Parrinello-Rahman barostat. Following equilibration, all positional restraints were removed, and production simulations were performed using a 2 fs integration time step. Covalent

bonds involving hydrogen atoms were constrained using the linear constraint solver (LINCS) algorithm. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with a real-space cutoff of 10 Å. Temperature and pressure were maintained at 310 K and 1 atm throughout the simulations using the velocity-rescale thermostat and isotropic Parrinello-Rahman pressure coupling, respectively. For each selected ternary complex, two independent production simulations of 100 ns each were performed using different initial velocity distributions to evaluate the reproducibility and robustness of the observed conformational behavior. The resulting trajectories were subsequently used for structural stability analysis, interfacial interaction assessment, and MM/PBSA binding free-energy calculations.

Trajectory Analysis, Conformational Dynamics, and Binding Free-Energy Calculations: All analyses were performed on the production trajectories obtained from the independent 100 ns molecular dynamics simulations. Prior to analysis, periodic boundary artifacts were removed, and trajectories were centered and aligned to the protein Cα atoms to eliminate rotational and translational motions.

Structural Stability Analysis: The overall stability and convergence of the Tau-PROTAC-CRBN ternary complexes were assessed by calculating the root mean square deviation (RMSD), and radius of gyration (Rg). RMSD was used to evaluate the structural stability of the complexes throughout the simulation, whereas Rg was monitored to assess compactness and global conformational changes during the simulations.

Hydrogen Bond occupancy Analysis: To investigate the role of PROTAC-mediated interactions in stabilizing the ternary complex, intermolecular hydrogen bonds were analyzed throughout the trajectories using the custom script. Hydrogen bonds were defined using geometric criteria consisting of a donor-acceptor distance ≤ 3.5 Å and a donor-hydrogen-acceptor angle $\geq 30°$. For each identified hydrogen bond, occupancy, lifetime distribution, and average persistence were calculated. In addition, hydrogen-bond existence maps and lifetime heatmaps were generated to visualize the temporal stability of key interactions. These analyses were used to identify persistent contacts contributing to ternary-complex stabilization.

Protein-PROTAC and Protein-Protein Contact Analysis: Intermolecular contacts between Tau, CRBN, and the PROTAC molecules were characterized using residue-contact frequency maps and interaction fingerprints. Contact occupancies were calculated for individual residues throughout the trajectories to identify residues contributing most strongly to ternary-complex stability. Persistent contacts observed across both simulation replicas were considered indicators of stable PROTAC-mediated protein association.

Solvent Accessible Surface Area Analysis: The solvent-accessible surface area (SASA) was calculated to evaluate protein surface burial and interface formation during the simulations. Changes in SASA were used to estimate the extent of interface stabilization between Tau and CRBN induced by the PROTAC molecules and to identify residues becoming buried upon ternary-complex formation.

Essential Dynamics and Free-Energy Landscape Analysis: Principal component analysis (PCA) was performed on Cα atomic coordinates to characterize dominant collective motions within the ternary complexes. The covariance matrix of atomic fluctuations was constructed and diagonalized to obtain the principal eigenvectors describing major conformational transitions. Trajectories were

projected onto the first two principal components (PC1 and PC2) and used to construct free-energy landscapes (FELs) according to:

$$\boldsymbol{F = -k_B T \ln P(PC1, PC2)} \quad \text{---} \quad \textbf{(7)}$$

where k_B is the Boltzmann constant, T is the simulation temperature, and P represents the probability density associated with each conformational state. The FELs were used to identify energetically favorable conformations and assess conformational convergence.

Dynamic Cross-Correlation Analysis: Dynamic cross-correlation matrices (DCCMs) were calculated to investigate correlated and anti-correlated motions within the ternary complexes. Correlation patterns were analyzed to determine whether PROTAC binding promoted coordinated motions between Tau and CRBN that could support productive ubiquitination.

Conformational Clustering: Trajectory frames were clustered using the GROMOS clustering algorithm with an RMSD cutoff of 2.0-3.0 Å. Representative structures from the most populated clusters were extracted and used to characterize dominant conformational states adopted by the ternary complexes during the simulations. Cluster populations were additionally used as a measure of conformational convergence and stability.

MM/PBSA Binding Free-Energy Calculations: Binding free energies were estimated using the **g_MMPBSA** package [35], [36]. Snapshots were extracted at regular intervals (every 100 ps) from the equilibrated portions of each trajectory and subjected to Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) calculations. Mean-binding free energies and standard deviations were calculated from all analyzed frames. To identify energetic hotspots contributing to complex stabilization, per-residue energy decomposition analyses were performed. Where applicable, energetic contributions for the binary interactions (Tau–PROTAC and CRBN–PROTAC) were compared with those of the complete ternary complex to evaluate the energetic contribution of PROTAC-mediated protein association.

Statistical Analysis: All reported MD-derived metrics were calculated independently for both simulation replicas (n = 2). Mean values and standard deviations were computed across replicas. For MM/PBSA calculations, confidence intervals were estimated by bootstrap resampling using 1000 iterations. Statistical comparisons between candidate PROTACs were performed using parametric or non-parametric tests as appropriate based on data distribution.

Software and scripts: Trajectory analyses were performed using **GROMACS**, **MDAnalysis**, and custom Python scripts. Molecular visualization was carried out using MOE v2024.6, SWISSpdb viewer v4.1.0, VMD v1.9.4a53 and PyMOL v4.60.

3. Results

3.1. Construction of a Combinatorial Tau-Targeting PROTAC Library

A virtual library of 16,400 Tau-targeting PROTACs was generated by systematically combining 41 curated Tau-binding ligands, 20 linker scaffolds, and 10 CRBN-recruiting ligand scaffolds with two alternative exit-vector variants, designated C4 and C5. The principal components of the combinatorial design are illustrated in Figure 1.

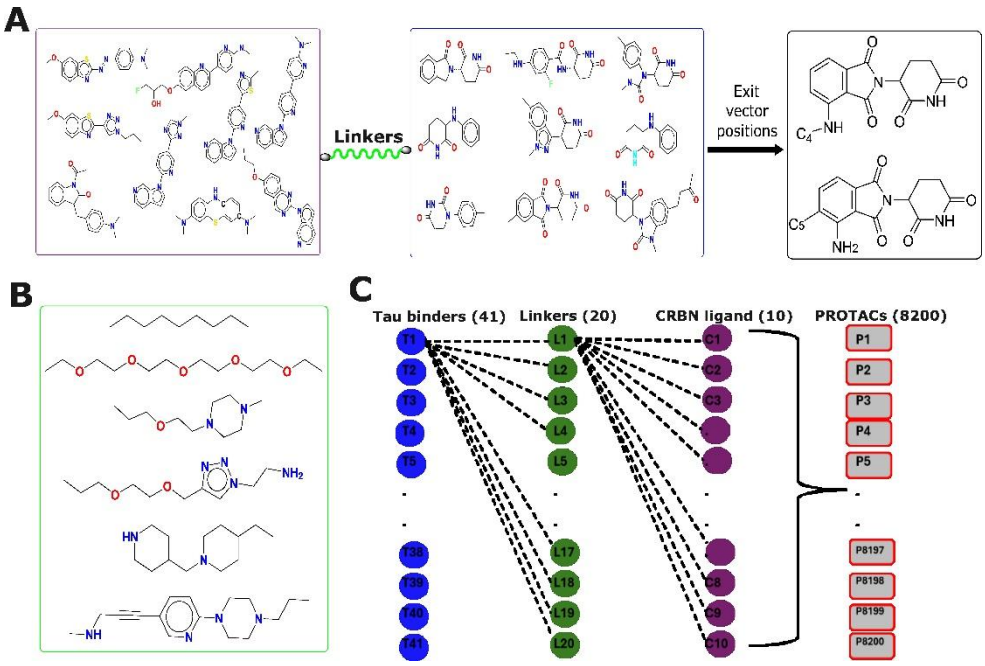


Figure 1. Combinatorial design and enumeration of the virtual Tau-targeting PROTAC library. (A) Representative Tau-binding and CRBN-recruiting ligands. C4 and C5 denote alternative linker-attachment positions on the CRBN-binding scaffolds. **(B)** Representative linkers varying in length, flexibility, cyclicity, and heteroatom content. **(C)** Enumeration of 41 Tau ligands, 20 linkers, and 10 CRBN recruiters generated 8,200 structures per attachment variant and 16,400 PROTACs in total. Complete set of E3 ligase ligands, linker moieties and Tau binders are provided in Supplementary File S1.

The 41 Tau-binding ligands were retrieved from BindingDB and curated to preserve scaffold diversity. Representative Tau binders and CRBN recruiters are shown in Figure 1A. Two attachment variants were defined on each CRBN-binding scaffold to account for exit-vector-dependent differences in the geometry of the assembled PROTACs. Linker diversity was introduced using 20 scaffolds that varied in length, flexibility, cyclicity, and heteroatom content, with representative structures shown in Figure 1B. Combinatorial enumeration of 41 Tau binders, 20 linkers, and 10 CRBN recruiters generated 8,200 structures for each attachment variant. Inclusion of both C4 and C5 variants yielded a total of 16,400 enumerated PROTAC structures, as represented by the theoretical design matrix in Figure 1C (Complete list of Tau binders, linkers and CRBN ligands provided in supplementary file 1).

3.2. Benchmarking and Weight-Sensitivity Analysis Support the Robustness of BT-MPO Framework

Before applying BT-MPO to the generated PROTAC library, we benchmarked the framework against published Tau-targeting and CNS-directed PROTACs and evaluated the sensitivity of compound rankings to the author-defined weights.

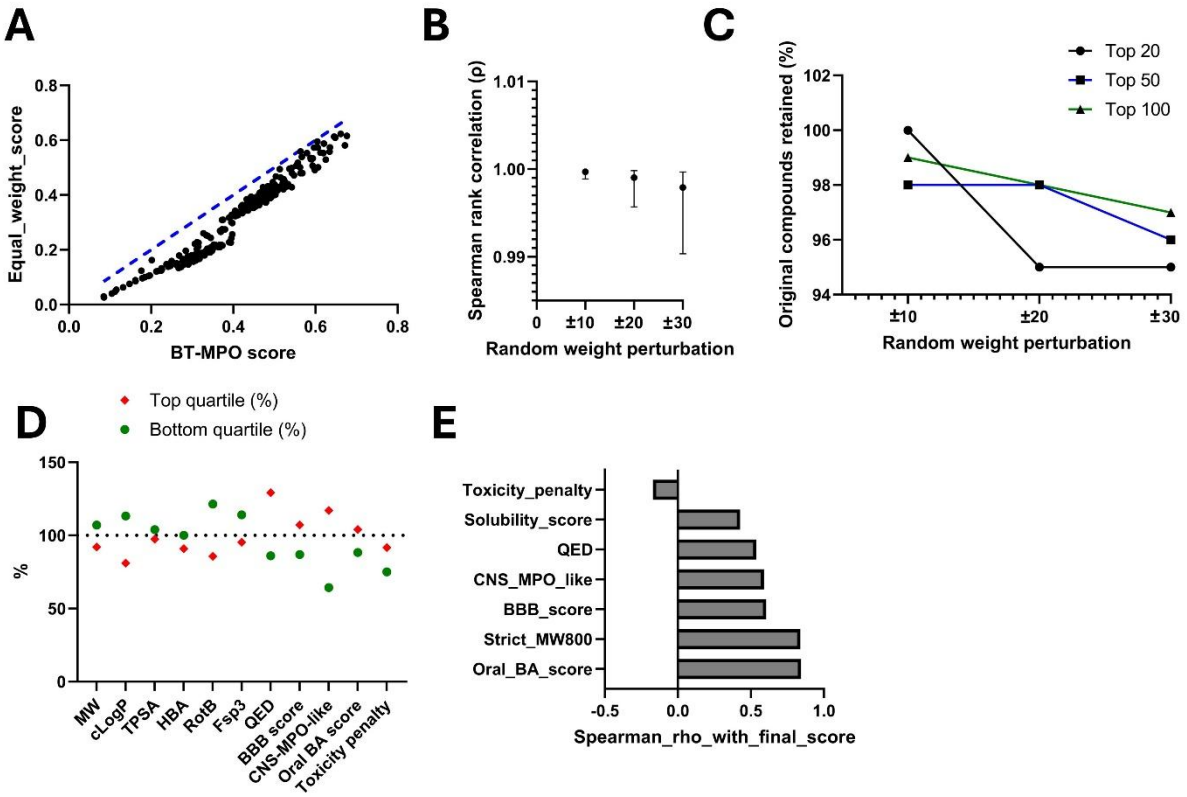


Figure 2. Robustness and physicochemical characteristics of the BT-MPO framework. (A) Relationship between the predefined BT-MPO score and an equal-weight comparator for 362 Tau-targeting PROTACs. **(B)** Spearman correlation with the original ranking following ±10%, ±20%, and ±30% Monte Carlo perturbation of the component weights, with 5,000 models evaluated per level. Points and error bars show medians and 2.5th–97.5th percentile intervals, respectively. **(C)** Retention of compounds from the original top-20, top-50, and top-100 sets following weight perturbation. **(D)** Relative physicochemical and calculated-property profiles of compounds in the upper and lower BT-MPO quartiles, normalized to the median of the complete Tau-targeting set. **(E)** Spearman correlations between BT-MPO and its component scores.

Monte Carlo perturbation confirmed that BT-MPO ranking was insensitive to moderate changes in the predefined weights. Median Spearman correlations with the original ranking were 0.9997, 0.9990, and 0.9979 following ±10%, ±20%, and ±30% perturbation, respectively; even at ±30%, the 2.5th percentile remained 0.9904 (Figure 2B). Median retention of the original top-50 compounds was 98%, 98%, and 96%, while top-100 retention was 99%, 98%, and 97%, respectively (Figure 2C). Thus, candidate ordering was not critically dependent on the precise numerical coefficients.

Upper-quartile BT-MPO compounds were smaller, less lipophilic, and less flexible than lower-quartile compounds, with lower MW, cLogP, TPSA, HBA count, and RotB count (Figure 2D). They also showed higher QED, heuristic BBB-suitability, modified CNS MPO-like, oral-developability, and solubility-desirability values. HBD count and approximate descriptor-derived logBB did not differ significantly between quartiles. Complete median values and statistical comparisons are provided in Supplementary Data File 2.

Component analysis showed the strongest associations with BT-MPO for the heuristic oral-developability index ($\rho = 0.844$) and MW compliance ($\rho = 0.840$), followed by the heuristic

BBB-suitability index ($\rho = 0.606$), modified CNS MPO-like score ($\rho = 0.591$), QED ($\rho = 0.538$), and solubility-desirability score ($\rho = 0.427$); the structural-alert and safety-proxy penalty was inversely associated with BT-MPO ($\rho = -0.170$; Figure 2E). Because these variables contributed directly or indirectly to the BT-MPO calculation, these associations describe the behavior of the scoring framework rather than provide independent validation.

The external CNS-directed PROTAC set had a comparable median BT-MPO score to the Tau-targeting set (0.393 versus 0.435; $P = 0.153$; Supplementary Figure S1A) but occupied a distinct physicochemical region. The reference compounds had higher TPSA, HBD count, and RotB count, together with lower approximate descriptor-derived logBB values and heuristic BBB-suitability indices (Supplementary Figure S1B–D and Supplementary Data File 3). Collectively, these analyses established the robustness of BT-MPO ranking and showed that reported CNS-directed PROTACs occupy heterogeneous physicochemical space. Because experimental brain exposure and degradation data were not uniformly available, the external comparison provided chemical-space context rather than validation of biological predictive performance. The BT-MPO framework was therefore applied to the generated library as a relative candidate-prioritization method.

3.3. Multi-Stage Stability and Pharmacokinetic Screening Prioritizes Tau-Targeting PROTACs

The 16,400-member library was subjected to sequential stability, pharmacokinetic, CNS-oriented, and chemical-space analyses. Combined plasma- and chemical-stability prediction retained 3,601 of 8,200 C4 PROTACs (43.9%) and 3,741 of 8,200 C5 PROTACs (45.6%), yielding 7,342 compounds for subsequent evaluation (Figure 3A). The similar retention rates indicated comparable overall predicted stability profiles for the two attachment points. Linker features showed substantial differences in predicted stability (Figure 3B). The highest stable proportions were observed for azo-containing linkers (62.5%), followed by pyridine and heteroaromatic linkers (54.3% each), urea linkers (53.8%), and triazole linkers (53.6%). Pyrazine-, alkyne-, and alkene-containing linkers showed lower stable proportions of 27.1%, 21.1%, and 6.7%, respectively, while no carbamate- or ester-containing constructs were classified as stable. Because the linker annotations overlapped, these values were interpreted as feature-level associations rather than independent effects of individual linker chemistry. Figure 3C shows the corresponding numbers of stable and unstable compounds associated with each feature and illustrates their unequal representation within the library. The ADMET platforms reduced the number of PROTAC libraries to 2911 for C4 and 3,534 for C5. Pearson-correlation analyses showed similar relationships in both libraries (Figure 3D and Supplementary Figure S2). cLogP was strongly inversely correlated with ESOL-predicted solubility (C4, $r = -0.943$; C5, $r = -0.947$), while TPSA was strongly inversely correlated with the approximate descriptor-derived logBB value (C4, $r = -0.931$; C5, $r = -0.930$). Increasing rotatable-bond count was associated with greater KierPhi and lower QED. Correlations between the BT-MPO score and its component variables were treated as internal checks of the predefined scoring framework rather than independent validation of CNS exposure or degradation activity.

None of the evaluated PROTACs simultaneously satisfied all eight stringent CNS-oriented reference criteria. Candidates were therefore ranked according to their overall BT-MPO profiles,

and the 50 highest-ranked compounds from each attachment-variant library were retained. Both selected subsets satisfied a median of 6 of 8 criteria (interquartile range: 5-6), corresponding to a median criteria coverage of 75.0%. Across the combined set, 8 compounds satisfied seven criteria, 47 satisfied six, 43 satisfied five, and 2 satisfied four. The median BT-MPO score was 0.688 for C4 candidates (interquartile range: 0.673–0.712; range: 0.659–0.781) and 0.661 for C5 candidates (interquartile range: 0.649–0.677; range: 0.642–0.721). All selected compounds met the MW, HBA, and synthetic-accessibility criteria, while 93% and 89% of the combined set met the TPSA and HBD criteria, respectively. The main predicted liabilities were insufficient ESOL-derived solubility, observed in 82% of both subsets; failure to meet the heuristic BBB-suitability criterion, observed in 74% of C4 and 78% of C5 candidates; and cLogP outside the proxy interval of 3.0–4.0, observed in 66% and 60%, respectively. Distribution histograms and selected molecules' PCA chemical space are shown in Supplementary Figure S3 (C4 for panel A and C5 for panel B). Both libraries displayed closely aligned, highly favorable property profiles, with median BBB scores of 0.80 (C4) and 0.78 (C5), molecular weights centered at 777.91 Da (C4) and 788.54 Da (C5), and narrow topological polar surface area (TPSA) distributions with medians of 143.53 Å² (C4) and 144.84 Å². Oral bioavailability scores skewed strongly toward high desirability, achieving median values of 0.85 for both series. The final BT-MPO scores exhibited characteristic bimodal distributions with medians of 0.45 (C4) and 0.47 (C5), while descriptor-based Principal Component Analysis (PCA) confirmed that selected candidate molecules (marked with 'x' markers) span across the optimal, high-desirability chemical space region (Supplementary Figure S3).

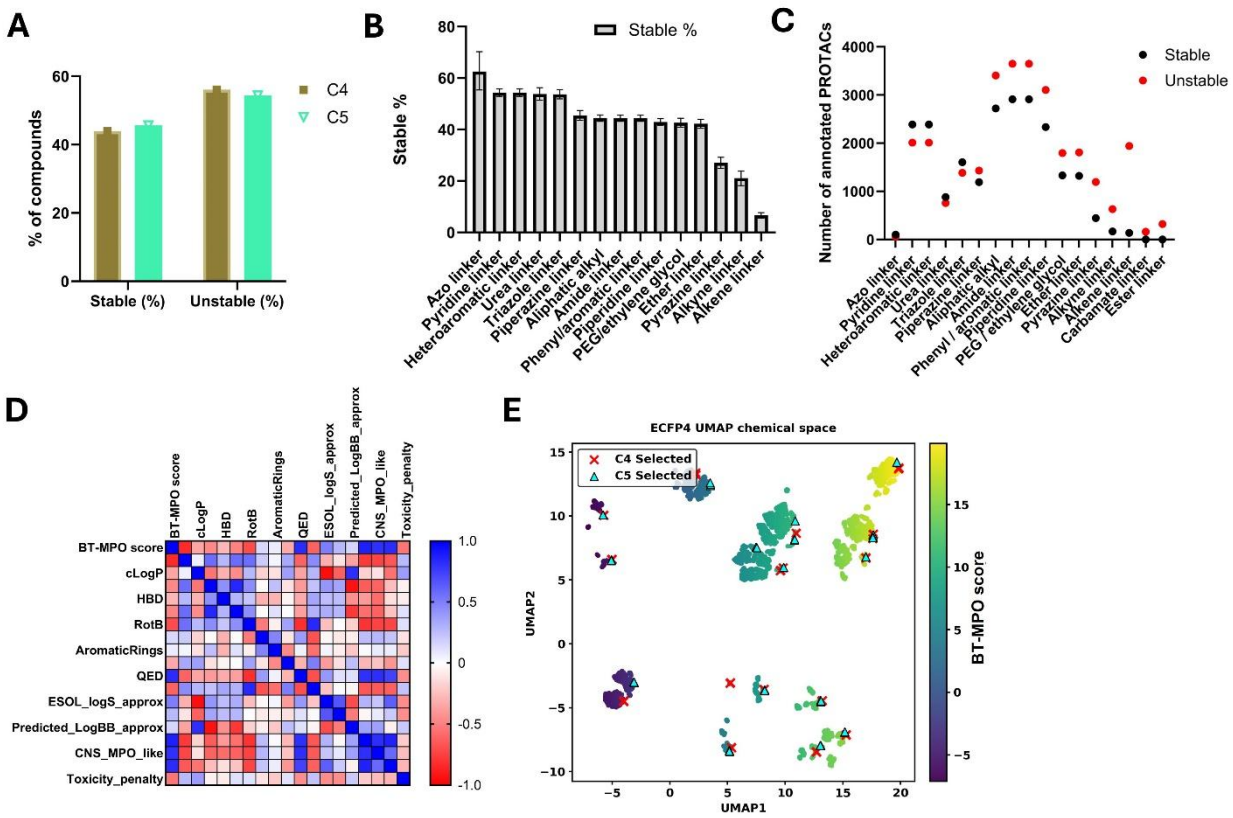


Figure 3. Multi-stage stability, linker-feature, pharmacokinetic, and chemical-space evaluation of the Tau-targeting PROTAC libraries. (A) Predicted plasma- and chemical-stability classifications for the C4

and C5 libraries. **(B)** Stable proportion within each overlapping linker-feature category. Points show percentages, whiskers indicate Wilson 95% confidence intervals, and labels show stable/total counts.(C) Numbers of stable and unstable compounds associated with each linker feature. Because feature annotations overlapped, compounds could contribute to multiple categories. **(D)** Pearson-correlation heatmap for physicochemical, ADMET, CNS-related, and composite-score variables in the complete-case C4 population ($n = 2{,}911$). Colors indicate correlation coefficients from -1 to +1; *P*values were adjusted using the Benjamini-Hochberg procedure. The corresponding C5 analysis ($n = 3{,}534$) is provided in Supplementary Figure S2, with complete statistics in Supplementary Table 1. **(E)** ECFP4-based UMAP representation of the combined C4 and C5 populations. Points are colored by BT-MPO score, and the 50 highest-ranked candidates from each library are overlaid using distinct symbols. HDBSCAN identified 15 regions, of which 14 contained at least one prioritized candidate from each attachment variant.

ECFP4-based UMAP analysis identified 15 fingerprint-defined regions within the common C4/C5 chemical space (Figure 3E). The selected C4 and C5 subsets each occupied 14 regions, corresponding to 93.3% regional coverage. Thus, prioritization reduced the original library to 100 best-balanced candidates, representing 0.61% of the initial design space, while retaining broad structural coverage. These compounds were subsequently advanced to Tau-PROTAC-CRBN ternary-complex prediction.

3.4. Linker Architecture Influences Predicted Tau-PROTAC-CRBN Ternary-Complex Formation

(The 100 candidates prioritized through stability, pharmacokinetic, BT-MPO, and chemical-space analyses were evaluated for formation of Tau-PROTAC-CRBN ternary complexes using the MOE Method 4B workflow. DCP values varied substantially among the modeled candidates, indicating that the Tau-binding ligand, linker architecture, CRBN-recruiting ligand, and attachment geometry collectively influenced the ability to generate consistently populated ternary-complex configurations (Figure 4A). The relationship between linear linker length and DCP was non-monotonic (Figure 4B). The highest DCP values were predominantly observed for PROTACs with linkers containing approximately 8-19 atoms, whereas shorter and longer linkers generally produced less populated ternary-complex solutions. However, compounds with similar linker lengths frequently exhibited markedly different DCP values, showing that linker length alone was insufficient to explain the modeled ternary-complex behavior. This effect was particularly evident for three representative PROTACs with comparable linear linker lengths but distinct linker compositions (Figure 4C). A flexible PEG-containing linker produced a DCP value of 462, whereas incorporation of a glycyl-containing unit reduced the DCP value to 6. Introduction of a triazole-containing element further reduced the DCP value to 3. Thus, these representative compounds showed more than a 70-fold difference in DCP despite having comparable nominal linker lengths, emphasizing the importance of linker flexibility, local geometry, and functional-group composition. Inspection of the highest-populated pose generated for the PEG-linked PROTAC showed a ternary arrangement that accommodated the Tau-binding and CRBN-recruiting moieties without obvious interfacial overlap in the selected model. The Tau residues involving Ser352, Gln351, and Lys353 were observed to interact with the PROTAC (Figure 4D and subset panel). The full 3D orientation of PEG linked PROTAC in interphase of CRBN and Tau show accessible lysine residues (Supplementary Figure S4A). These interactions describe the selected static pose and were used to identify candidate interfacial contacts for subsequent MD

assessment. The glycyl-linker PROTAC showed a rich environment of lysines near the tau binding scaffold; however, the ternary complex shows restrained orientation of PROTAC into protein-protein interphase explains the reduced DCP (Supplementary Figure S4B and S4C). The ternary complex generated with the triazole-linker PROTAC shows an extended PROTAC stretch across both tau and CRBN, suggesting over-flexibility and hence reduced ternary complex formation (Figure 4E). Superposition of the triazole-containing PROTAC, glycyl-linker PROTAC onto the PEG-linked PROTAC conformations extracted from ternary complex revealed a change in the relative orientation in proteins interphase (Figure 4F).

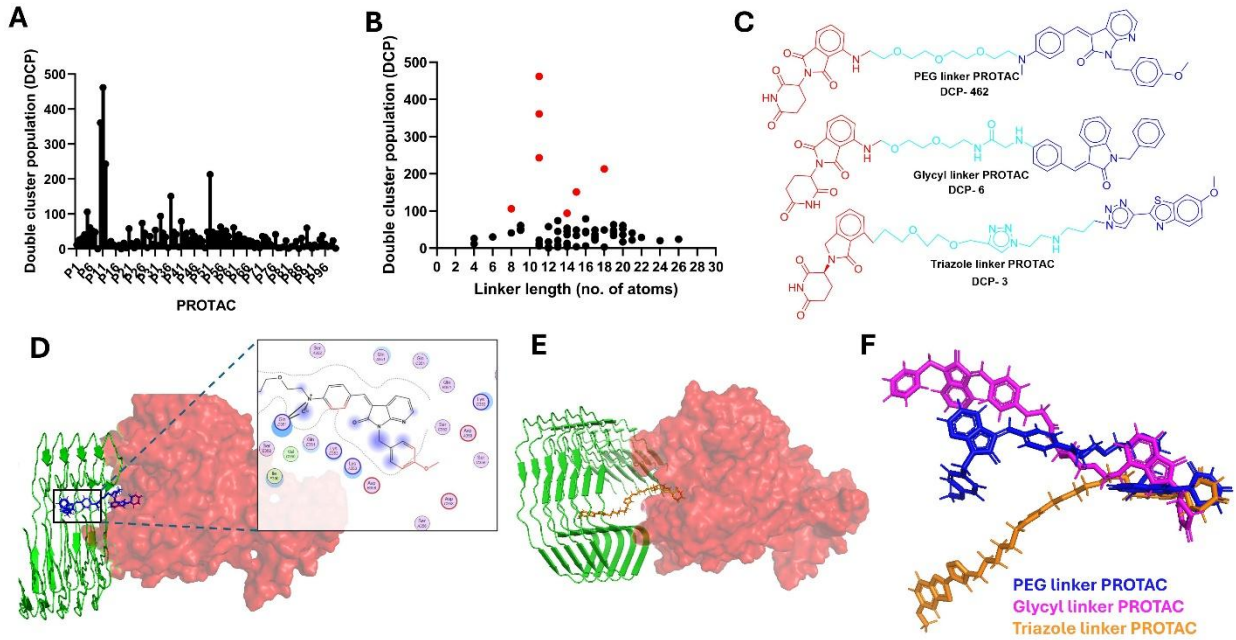


Figure 4. Linker-dependent differences in modeled Tau-PROTAC-CRBN ternary-complex formation. **(A)** Double Cluster Population (DCP) distribution for prioritized PROTACs modeled using MOE Method 4B. **(B)** Relationship between linker length and DCP. **(C)** Representative PEG-, glycyl-, and triazole-linked PROTACs with DCP values of 462, 6, and 3, respectively. **(D, E)** Representative modeled complexes of the PEG- and triazole-linked PROTACs, showing interfacial contacts and proximal Tau lysines. **(F)** Structural superposition illustrating linker-dependent differences in Tau–CRBN orientation. DCP, interfacial geometry, and solvent-exposed Tau lysines within 10 Å of the E3-facing region guided selection of five candidates for molecular-dynamics simulations. Additional interaction diagrams are provided in Supplementary Figure S4.

The triazole-containing PROTAC produced the lowest DCP value among the three representative molecules. The selected pose displayed less favorable interfacial accommodation than the PEG-linked complex, providing a structural hypothesis for the lower population of recurrent ternary-complex solutions. Structural superposition of the PEG, glycyl and triazole PROTAC orientations extracted from the ternary complex were shown (Figure 4F). The modeled poses were additionally examined for accessible Tau lysine residues located within 10 Å of the predefined E3-facing region. Candidates differed in both DCP and proximal-lysine coverage, indicating that a highly populated ternary-complex solution did not necessarily provide the same putative ubiquitination geometry across all compounds (Supplementary Figure S4). Lysine proximity was therefore used as a structural-prioritization criterion rather than as evidence of confirmed ubiquitination. Based on the combined consideration of DCP, ternary-complex geometry, interfacial contacts, and proximal

solvent-exposed lysine residues, five candidate PROTACs were selected for two independent 100-ns MD simulations. This selection retained compounds with favorable modeled ternary complex populations while also considering geometries potentially compatible with Tau ubiquitination.

3.5. Molecular Dynamics Simulations Reveal Candidate-Dependent Ternary Complex Stability

The PROTACs selected from the MOE ternary-complex analysis, hereafter designated PROTACs 1-5, were subjected to two independent 100-ns all-atom MD simulations to evaluate the stability and conformational behavior of the modeled Tau-PROTAC-CRBN complexes. Backbone RMSD analysis revealed candidate-dependent differences in ternary-complex stability (Figure 5A). PROTACs 2-5 approached relatively stable RMSD plateaus after approximately 40 ns, with limited frame-to-frame variation during the later portions of the trajectories. In contrast, the complex containing PROTAC 1 showed greater and more persistent fluctuations, with RMSD values distributed broadly between approximately 0.5 and 1.0 nm. These observations suggested that the modeled complex containing PROTAC 1 underwent larger conformational rearrangements than those formed by PROTACs 2-5. The complex containing PROTAC 5 initially exhibited an increase in RMSD, followed by a lower and more stable plateau. A similar pattern was observed in the radius of gyration analysis (Figure 5B), in which the Rg increased transiently from approximately 3.2 to 3.3 nm before decreasing to approximately 3.1 nm and remaining comparatively stable. This behavior was consistent with an initial structural rearrangement followed by formation of a more compact conformational ensemble. PROTACs 2-5 generally maintained stable or modestly decreasing Rg values, whereas the greater RMSD variation observed for PROTAC 1 indicated comparatively lower conformational stability. These analyses support differences in global compactness and stability but do not, by themselves, establish tighter binding or direct movement of Tau and CRBN toward one another.

Protein-ligand hydrogen-bond analysis was used to characterize the interactions associated with PROTAC 5 (Figure 5C). Tau residues Lys420 showed the highest hydrogen-bond occupancy at 35.3%, followed by Trp345 at 23.6%, Lys497 at 17.1%, Asn296 at 15.8%, and Lys574 at 10.7%. Other identified hydrogen bonds had occupancies below 10%. These results identified Lys420 and Trp345 as the most persistent hydrogen-bonding residues in the modeled PROTAC 5 complex. Contact-frequency analysis using a 4.5-Å cutoff further showed recurrent contacts between PROTAC 5 and residues including Trp345, Asn296, Lys420, Gln495, Ser496, Lys497, Gln726, and other neighboring residues (Figure 5D). Several contacts persisted throughout most or all of the analyzed trajectory frames, indicating that PROTAC 5 remained associated with a relatively consistent interaction environment during the simulations. The remaining candidates showed distinct hydrogen-bonding profiles. PROTAC 1 interacted most persistently with Trp319, with an occupancy of 51.3% (Supplementary Figure S5A), whereas His298 was the dominant hydrogen-bonding residue for PROTAC 2 (Supplementary Figure S5B), with an occupancy of 48.9%. PROTAC 3 formed a broader hydrogen-bonding network, including interactions with Lys204 and Ser126 at occupancies above 20% and with Ser203 and Ser357 at occupancies above 30% (Supplementary Figure S5C). In contrast, PROTAC 4 showed comparatively low hydrogen-bond occupancies, with its most persistent interaction involving Trp345 at 8.1% (Supplementary Figure S5D). Despite these differences, contact-frequency maps identified several recurrent residues across multiple complexes, including Trp345, Gln495, Lys728, and Trp325, suggesting partially shared interaction environments among the modeled candidates.

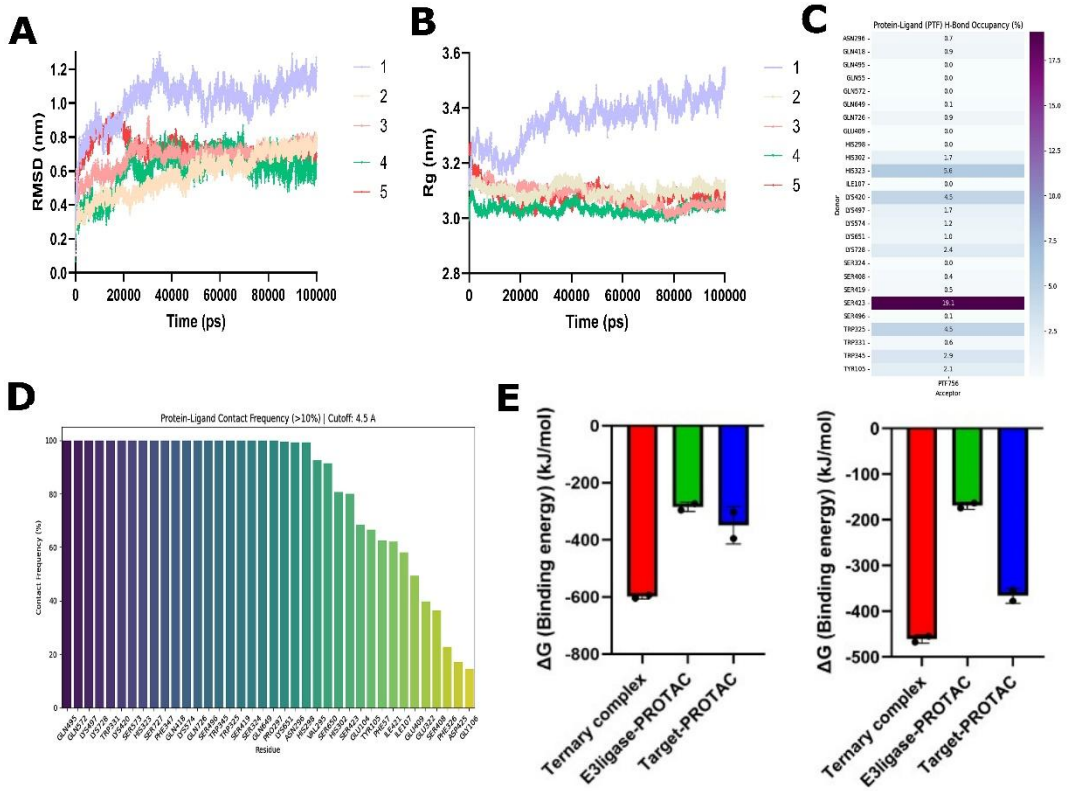


Figure 5. Molecular-dynamics and MM/PBSA analyses reveal candidate-dependent stability of modeled Tau–PROTAC–CRBN ternary complexes. (A, B) Backbone RMSD and radius of gyration, respectively, for complexes containing PROTACs 1-5 during two independent 100-ns simulations. **(C, D)** Hydrogen-bond occupancy and protein-contact frequency for PROTAC 5; contacts were defined using a 4.5 Å cutoff. **(E)** MM/PBSA energetic distributions for the ternary and binary systems containing

PROTACs 5 and 3. Values are reported as mean ± SD and interpreted comparatively rather than as absolute binding free energy. Corresponding interaction analyses for PROTACs 1–4 and MM/PBSA profiles for PROTACs 1, 2, and 4 are provided in Supplementary Figures S5 and S6, respectively.

MM/PBSA calculations were used to compare the relative energy profiles of the modeled complexes (Figure 5E). PROTAC 5 showed a mean ternary-complex MM/PBSA energy of -604 ± 34kcal/mol, which was more negative than the corresponding binary-complex estimates. This energetic profile was consistent with favorable stabilization of the modeled ternary assembly, although the absolute MM/PBSA value should be interpreted comparatively rather than as an experimental binding free energy (Figure 5E, left panel). For PROTAC 3, the mean ternary-complex energy was -467.5 ± 34kcal/mol, whereas the Tau-PROTAC binary interaction yielded -377.8 ± 21kcal/mol (Figure 5E, right panel). Although the ternary estimate was more negative than the reported Tau-PROTAC binary estimate, the difference was smaller than that observed for PROTAC 5, indicating a comparatively more limited energetic gain upon formation of the modeled ternary assembly. For PROTAC 3, the mean ternary-complex MM/PBSA estimate was -467.5 ± 34kcal/mol, compared with -377.8 ± 21kcal/mol for the Tau–PROTAC binary complex. The more negative ternary estimate was consistent with comparatively favorable modeled ternary association, although the energetic difference was smaller than that observed for PROTAC 5. . PROTAC 1 showed broadly similar MM/PBSA energy estimates for the ternary and binary systems, suggesting limited energetic differentiation among the modeled states (Supplementary Figure S6A). PROTACs 2 and 4 showed more negative ternary-complex energies than their corresponding binary estimates, consistent with comparatively favorable modeled ternary association (Supplementary Figure S6B and S6C). Together, the RMSD, Rg, interaction-occupancy, contact-frequency, and MM/PBSA analyses identified different stability profiles among the five modeled ternary complexes. PROTAC 5 showed the clearest combination of late-trajectory RMSD stabilization, compactness, persistent protein–ligand contacts, and favorable relative MM/PBSA energetics. PROTACs 2 and 4 also maintained comparatively stable ternary-complex trajectories, whereas PROTAC 1 showed greater conformational variability. Together, these complementary analyses identified the most structurally favorable modeled ternary complexes and supported the selection of leading candidates for experimental assessment of ternary-complex formation, ubiquitination, Tau degradation, and brain exposure.

3.6. Free-Energy Landscapes and Correlated Motions Reveal Distinct Ternary-Complex Dynamics

Principal-component-based free-energy landscapes and dynamic cross-correlation matrices were used to compare the conformational behavior of Tau-PROTAC-CRBN complexes containing PROTACs 1-5 (Figure 6A-E). The complexes containing PROTACs 1, 3, and 4 displayed broad, multi-basin landscapes spanning a wide region of principal-component space, with PC1 extending approximately from -150 to +100 (left panels of Figure 6A, 6C, and 6D respectively). These distributions indicated sampling of multiple conformational substates rather than concentration within a single dominant basin. The PROTAC 2 complex showed a more fragmented landscape with several separated local minima, indicating comparatively heterogeneous conformational sampling (Left panel of Figure 6 B).

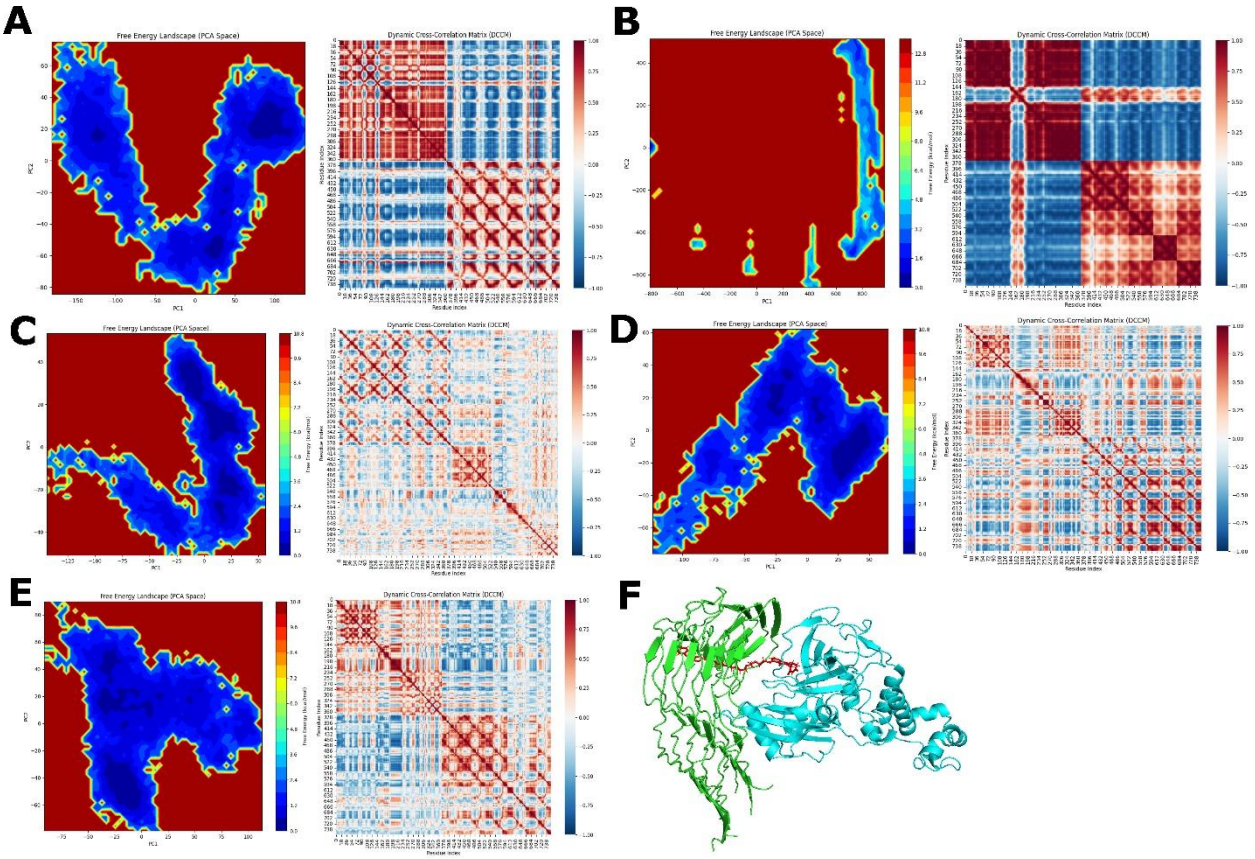


Figure 6. Free-energy landscapes and dynamic cross-correlation analyses of the modeled Tau-PROTAC-CRBN ternary complexes. (A-E) Free-energy landscapes and corresponding dynamic cross-correlation matrices for complexes containing PROTACs 1-5 respectively. Left panels show trajectories projected onto PC1 and PC2, with relative free energy calculated as $F = -k_{\mathrm{B}}T \ln P\,(P, C1PC2)$; lower-energy regions represent more frequently sampled states. The asterisk indicates the PROTAC 5 global minimum used for representative-structure extraction. Right panels show residue cross-correlations from -1 to +1, with red indicating correlated motion, blue indicating anticorrelated motion, and values near zero indicating weak correlation. Residues 1-380 and 381-760 correspond to CRBN and Tau, respectively, and off-diagonal regions represent interprotein coupling. **(F)** Representative structure from the principal PROTAC 5 low-energy basin, with Tau, CRBN, and PROTAC 5 shown as green ribbons, cyan ribbons, and red sticks, respectively.

In contrast, the PROTAC 5 complex exhibited a more localized free-energy landscape dominated by a principal low-energy basin near PC1 ≈ -15 and PC2 ≈ -35 (left panel of Figure 6E). The

concentration of trajectory frames within this restricted region indicated reduced conformational heterogeneity relative to PROTACs 1-4. This finding was consistent with the late-trajectory RMSD stabilization, reduction in radius of gyration to approximately 3.1 nm, persistent protein-ligand contacts, and the comparatively favorable MM/PBSA energetic profile observed for PROTAC 5.

Dynamic cross-correlation matrix analysis revealed candidate-dependent patterns of collective motion across the modeled ternary complexes (right panels in Figure 6A-E). The PROTAC 1 complex showed strong intraprotein correlations within Tau and the E3 ligase, together with pronounced interprotein anticorrelation, consistent with opposing, hinge-like motions between the two proteins (Figure 6A, right panel). PROTAC 2 displayed a similar pattern, although E3 ligase residues 144-180 showed distinct correlations with the tau region and the PROTAC, identifying this segment as a potential dynamic coupling region (Figure 6B, right panel). In contrast, the PROTAC 3 complex showed predominantly weak interprotein correlations, indicating comparatively independent Tau and E3-ligase motions despite preserved local intraprotein correlations (Figure 6C, right panel). PROTAC 4 exhibited a moderately coupled, multidomain pattern, with localized correlations connecting E3 ligase regions spanning residues 180-234 and 306-360 to the Tau and PROTAC (Figure 6D, right panel). The PROTAC 5 complex retained strong intraprotein correlations and showed distinct localized correlated and anticorrelated motions across the Tau-E3 ligase interface, together with selective PROTAC-associated correlation tracks connecting regions of both proteins (Figure 6E, right panel). Collectively, these patterns indicate that PROTAC 5 supported a comparatively coordinated but dynamically adaptable ternary-complex ensemble, consistent with its favorable conformational profile in the preceding analyses. Further, a representative structure extracted from the principal low-energy basin of the PROTAC 5 landscape showed Tau and CRBN closely associated around the centrally positioned PROTAC, without obvious interprotein overlapping upon visual inspection (Figure 6F). Collectively, the localized basin near PC1 ≈ -15 and PC2 ≈ -35, reduced interprotein anticorrelation, and strong intra-protein correlations supported the comparatively favorable modeled stability profile of the PROTAC 5 complex. Further experimental studies are needed to determine whether these computational observations translate into productive ternary-complex formation, ubiquitination, or Tau degradation.

4. Discussion:

Tau-targeted degradation offers an effective strategy to conventional inhibition because it seeks to remove intracellular disease-associated protein rather than modulate a single Tau interaction or upstream pathway. Previous studies have demonstrated Tau degradation in cellular, patient-derived neuronal, and animal models [7], [8], [9], [10], [11], [12]. The present study extends these advances by introducing a systematic computational framework that evaluates a large combinatorial Tau-targeting PROTAC library across stability, CNS-oriented properties, chemical diversity, ternary-complex formation, and conformational dynamics. This integrated strategy enabled the identification of structurally diverse, high-priority candidates before resource-intensive synthesis and experimental testing.

The BT-MPO framework was developed to address the competing property requirements of CNS-directed PROTACs. Molecular size, polarity, solubility, flexibility, and hydrogen-bonding capacity present challenges for PROTAC permeability and pharmacokinetics because these molecules

frequently occupy beyond-Rule-of-Five chemical space [3] . Brain-directed degraders face the additional requirement of achieving sufficient CNS exposure while retaining productive target and E3-ligase engagement [37], [38]. In this context, the absence of compounds satisfying every stringent reference criterion reflects the difficulty of simultaneously optimizing these properties rather than a limitation of relative multiparameter prioritization. The stability of BT-MPO rankings under equal weighting and Monte Carlo perturbation supports the robustness of candidate ranking, although it does not establish that the author-defined coefficients are uniquely optimal. Comparison with published Tau-targeting and CNS-directed PROTACs further showed that reported degraders occupy heterogeneous physicochemical space, supporting the use of continuous multiparameter ranking rather than conventional small-molecule thresholds alone.

The present workflow is consistent with the broader movement toward integrated computational PROTAC design. Deep-learning approaches have been used to accelerate rational PROTAC design by jointly considering target proteins, E3 ligases, and linker generation, mean-while AI-supported workflows have combined molecular generation, structural alignment, molecular dynamics, and free-energy calculations for candidate evaluation [39] . Similarly, computational reviews have highlighted the growing integration of chemoinformatics, ternary-complex modeling, molecular dynamics, machine learning, and ADMET prediction in PROTAC discovery [40], [41]. The present study adds a CNS-oriented dimension to these approaches by combining chemical-space preservation and BT-MPO ranking with structural evaluation of Tau-PROTAC-CRBN assemblies. The retention of broad ECFP4-based regional coverage indicates that the ranking process did not restrict candidate selection to a single closely related chemotype, which is advantageous when computational models may perform differently across large and flexible degrader structures.

The structural analyses reinforce the importance of evaluating the complete PROTAC architecture rather than considering the target ligand, E3 recruiter, or linker length independently. Previous linker studies have shown that linker composition and attachment geometry can markedly influence degrader activity and physicochemical behavior [42]. Computational ensemble approaches such as SILCS-xTAC likewise integrate geometric, energetic, and physicochemical features to evaluate alternative ternary-complex configurations [43]. Comparative benchmarking of PRosettaC and AlphaFold3 has further demonstrated that chemically defined restraints, linker sampling, accessory proteins, and protein flexibility influence ternary-complex prediction accuracy [44]. These studies support the conclusion that linker length alone is insufficient to predict favorable Tau–PROTAC–CRBN assembly. Notably, triazole-containing PROTACs showed a relatively high predicted stable proportion across the library, whereas the representative triazole-linked complex exhibited a low DCP value. This apparent divergence reflects the distinct properties captured by the two analyses: stability predictions estimate compound integrity under plasma and storage conditions, whereas DCP measures the recurrence of modeled ternary-complex configurations. The conformational restriction introduced by a triazole unit may enhance chemical robustness while limiting the linker torsions and protein orientations required for favorable ternary-complex assembly. This interpretation is consistent with previous studies showing that linker composition, rigidity, and attachment geometry independently influence PROTAC developability and ternary-complex behavior [42], [45]. Thus, linker optimization requires balancing molecular stability with the conformational adaptability needed for ternary-complex

formation. Convergence among ternary-complex geometry, interaction persistence, relative energetic estimates, conformational landscapes, and correlated motions subsequently provided a complementary basis for prioritizing PROTAC 5 without reliance on a single computational metric.

The present framework provides a reproducible route for reducing a large Tau-targeting PROTAC library to a focused and structurally diverse candidate set. Its principal strength lies in combining molecular-property assessment with ensemble-based structural evaluation, thereby addressing both CNS-oriented developability and ternary-complex compatibility. The custom BT-MPO components remain relative ranking indices, and degradation or brain-exposure measurements were not uniformly available across benchmark libraries. These limitations define clear opportunities for further development in the field. Prospective evaluation of the leading candidates can establish relationships among BT-MPO ranking, ternary-complex formation, Tau ubiquitination, cellular degradation, neuronal activity, and brain exposure. Integration of these experimental outcomes into future iterations could progressively calibrate the framework and improve its predictive relevance for CNS-directed Tau degrader development.

Author Contributions

B.S. conceived the study, developed the methodology, performed the computational analyses, prepared the visualizations, and drafted the manuscript. B.W. reviewed and edited the manuscript. L.T. supervised the study and reviewed and edited the manuscript. L.W. conceived the project and reviewed and edited the manuscript. All authors read and approved of the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by grants from Margaretha af Ugglas Foundation, Swedish Research Council (2024-03573), the private initiative “Innovative Ways to Fight Alzheimer's Disease – Leif Lundblad Family and Others,” Swedish Brain Foundation (Hjärnfonden), Swedish Alzheimer Foundation, Petrus and Augusta Hedlunds Foundation, Tore Nilssons Foundation for Medical Research, Maja & J.P Åhlén Foundation, Gamla Tjänarinnor Foundation, Gun and Bertil Stohnes Foundation, Strategic Research Program in Neuroscience (StratNeuro) funding for postdoctoral researchers at Karolinska Institutet, Foundation for Geriatric Diseases at Karolinska Institutet, and Karolinska Institutet Research Grants.

Computations were performed at NSC Tetralith/Dardel provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), funded by the Swedish Research Council.

Data availability

The scripts and other files are provided as supplementary files.

References:

[1] J. Liu *et al.*, “PROTACs: a novel strategy for cancer therapy,” in *Seminars in cancer biology*, Elsevier, 2020, pp. 171–179.

[2] Z. Liu *et al.*, “An overview of PROTACs: a promising drug discovery paradigm,” *Molecular biomedicine*, vol. 3, no. 1, p. 46, 2022.

[3] S. D. Edmondson, B. Yang, and C. Fallan, “Proteolysis targeting chimeras (PROTACs) in ‘beyond rule-of-five’chemical space: Recent progress and future challenges,” *Bioorg. Med. Chem. Lett.*, vol. 29, no. 13, pp. 1555–1564, 2019.

[4] T. Tashima, “Proteolysis-targeting chimera (PROTAC) delivery into the brain across the blood-brain barrier,” *Antibodies*, vol. 12, no. 3, p. 43, 2023.

[5] T. T. Wager, X. Hou, P. R. Verhoest, and A. Villalobos, “Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties,” *ACS Chem. Neurosci.*, vol. 1, no. 6, p. 435, 2010.

[6] J. Gu and F. Liu, “Tau in Alzheimer’s disease: pathological alterations and an attractive therapeutic target,” *Curr. Med. Sci.*, vol. 40, no. 6, pp. 1009–1021, 2020.

[7] T.-T. Chu *et al.*, “Specific knockdown of endogenous tau protein by peptide-directed ubiquitin-proteasome degradation,” *Cell Chem. Biol.*, vol. 23, no. 4, pp. 453–461, 2016.

[8] M. Lu *et al.*, “Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway,” *Eur. J. Med. Chem.*, vol. 146, pp. 251–259, 2018.

[9] M. C. Silva *et al.*, “Targeted degradation of aberrant tau in frontotemporal dementia patient-derived neuronal cell models,” *Elife*, vol. 8, p. e45457, 2019.

[10] M. C. Silva *et al.*, “Discovery and optimization of tau targeted protein degraders enabled by patient induced pluripotent stem cells-derived neuronal models of tauopathy,” *Front. Cell. Neurosci.*, vol. 16, p. 801179, 2022.

[11] W. Wang *et al.*, “A novel small-molecule PROTAC selectively promotes tau clearance to improve cognitive functions in Alzheimer-like models,” *Theranostics*, vol. 11, no. 11, p. 5279, 2021.

[12] D. Yao *et al.*, “Selective degradation of hyperphosphorylated tau by proteolysis-targeting chimeras ameliorates cognitive function in Alzheimer’s disease model mice,” *Front. Pharmacol.*, vol. 15, p. 1351792, 2024.

[13] M. L. Drummond and C. I. Williams, “In silico modeling of PROTAC-mediated ternary complexes: validation and application,” *J. Chem. Inf. Model.*, vol. 59, no. 4, pp. 1634–1644, 2019.

[14] M. L. Drummond, A. Henry, H. Li, and C. I. Williams, “Improved accuracy for modeling PROTAC-mediated ternary complex formation and targeted protein degradation via new in silico methodologies,” *J. Chem. Inf. Model.*, vol. 60, no. 10, pp. 5234–5254, 2020.

[15] H. Kumar and M. E. Sobhia, “Interplay of PROTAC complex dynamics for undruggable targets: insights into ternary complex behavior and linker design,” *ACS Med. Chem. Lett.*, vol. 15, no. 8, p. 1306, 2024.

[16] M. S. Martinez-Cortés, C. A. Velázquez-Martínez, and J. L. Medina-Franco, “Advances in computational methods for PROTAC drug discovery,” *Drug Discov. Today*, p. 104627, 2026.

[17] W. D. Jang, J. Jang, J. S. Song, S. Ahn, and K.-S. Oh, “PredPS: Attention-based graph neural network for predicting stability of compounds in human plasma,” 2023.

[18] X. Li, X. Yan, Q. Gu, H. Zhou, D. Wu, and J. Xu, “Deepchemstable: chemical stability prediction with an attention-based graph convolution network,” *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 1044–1049, 2019.

[19] Y. Myung, A. G. C. de Sá, and D. B. Ascher, “Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction,” *Nucleic Acids Res.*, vol. 52, no. W1, pp. W469–W475, 2024.

[20] A. Daina, O. Michielin, and V. Zoete, “SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules,” *Sci. Rep.*, vol. 7, no. 1, p. 42717, 2017.

[21] L. Fu *et al.*, “ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support,” *Nucleic Acids Res.*, vol. 52, no. W1, pp. W422–W431, 2024.

[22] T. T. Wager, X. Hou, P. R. Verhoest, and A. Villalobos, “Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties,” *ACS Chem. Neurosci.*, vol. 1, no. 6, p. 435, 2010.

[23] M. Gupta, H. J. Lee, C. J. Barden, and D. F. Weaver, “The blood–brain barrier (BBB) score,” *J. Med. Chem.*, vol. 62, no. 21, pp. 9824–9836, 2019.

[24] J. L. Baylon *et al.*, “New multiparameter index is a strong predictor of oral bioavailability for heterobifunctional degraders,” *ACS Med. Chem. Lett.*, vol. 16, no. 6, p. 1108, 2025.

[25] G. R. Bickerton, G. V Paolini, J. Besnard, S. Muresan, and A. L. Hopkins, “Quantifying the chemical beauty of drugs,” *Nat. Chem.*, vol. 4, no. 2, pp. 90–98, 2012.

[26] J. B. Baell and G. A. Holloway, “New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays,” *J. Med. Chem.*, vol. 53, no. 7, pp. 2719–2740, 2010.

[27] A. F. Stepan *et al.*, “Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States,” *Chem. Res. Toxicol.*, vol. 24, no. 9, pp. 1345–1410, 2011.

[28] I. Sushko, E. Salmina, V. A. Potemkin, G. Poda, and I. V Tetko, “ToxAlerts: a web server of structural alerts for toxic chemicals and compounds with potential adverse reactions,” *J. Chem. Inf. Model.*, vol. 52, no. 8, p. 2310, 2012.

[29] C. Limban *et al.*, “The use of structural alerts to avoid the toxicity of pharmaceuticals,” *Toxicol. Rep.*, vol. 5, pp. 943–953, 2018.

[30] J. Ge *et al.*, “PROTAC-DB 3.0: an updated database of PROTACs with extended pharmacokinetic parameters,” *Nucleic Acids Res.*, vol. 53, no. D1, pp. D1510–D1515, 2025.

[31] D. Rogers and M. Hahn, “Extended-connectivity fingerprints,” *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010.

[32] M. L. Drummond, A. Henry, H. Li, and C. I. Williams, “Improved accuracy for modeling PROTAC-mediated ternary complex formation and targeted protein degradation via new in silico methodologies,” *J. Chem. Inf. Model.*, vol. 60, no. 10, pp. 5234–5254, 2020.

[33] M. J. Abraham *et al.*, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1, pp. 19–25, 2015.

[34] N. Schmid *et al.*, “Definition and testing of the GROMOS force-field versions 54A7 and 54B7,” *European biophysics journal*, vol. 40, no. 7, pp. 843–856, 2011.

[35] R. Kumari, R. Kumar, O. S. D. D. Consortium, and A. Lynn, “g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA calculations,” *J. Chem. Inf. Model.*, vol. 54, no. 7, pp. 1951–1962, 2014.

[36] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, “Electrostatics of nanosystems: application to microtubules and the ribosome,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 18, pp. 10037–10041, 2001.

[37] T. Tashima, “Proteolysis-targeting chimera (PROTAC) delivery into the brain across the blood-brain barrier,” *Antibodies*, vol. 12, no. 3, p. 43, 2023.

[38] X. Li *et al.*, “Engineering brain-penetrant PROTACs: Bridging molecular design and CNS delivery,” *Adv. Drug Deliv. Rev.*, vol. 234, no. 11587, p. 115876, 2026.

[39] C.-L. Chou, C.-T. Lin, C.-T. Kao, and C.-C. Lin, “A novel rational PROTACs design and validation via AI-driven drug design approach,” *ACS Omega*, vol. 9, no. 37, p. 38371, 2024.

[40] M. S. Martinez-Cortés, C. A. Velázquez-Martínez, and J. L. Medina-Franco, “Advances in computational methods for PROTAC drug discovery,” *Drug Discov. Today*, p. 104627, 2026.

[41] S. Zheng *et al.*, “Accelerated rational PROTAC design via deep learning and molecular simulations,” *Nat. Mach. Intell.*, vol. 4, no. 9, pp. 739–748, 2022.

[42] R. I. Troup, C. Fallan, and M. G. J. Baud, “Current strategies for the design of PROTAC linkers: a critical review,” *Explor. Target. Antitumor Ther.*, vol. 1, no. 5, p. 273, 2020.

[43] E. B. Nordquist, M. Zhao, W. Yu, and A. D. MacKerell Jr, “Computational modeling of PROTAC ternary complexes as ensembles using SILCS-xTAC,” *J. Chem. Inf. Model.*, vol. 66, no. 1, pp. 701–715, 2026.

[44] J. M. Schulz, S. I. Schürer, R. C. Reynolds, and S. C. Schürer, “PRosettaC outperforms AlphaFold3 for modeling PROTAC ternary complexes,” *Sci. Rep.*, vol. 15, no. 1, p. 37620, 2025.

[45] R. P. Wurz *et al.*, “Affinity and cooperativity modulate ternary complex formation to drive targeted protein degradation,” *Nat. Commun.*, vol. 14, no. 1, p. 4177, 2023.

Supplementary Files

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

- Supplementarydata.zip