Supplementary Information for UniLingo3DMol: a unified 3D molecular language model for *de novo* and fragment-based drug design

Han Wang[1,2†]*, Guanglong Sun[2†], Bowen Zhang[3†], Yang Wang[2†], Bin Xi[1,5†], Wenbiao Zhou[4†], Minjian Yang[1,2], Chuanyu Liu[3], Yuyang Ge[2], Fan Fan[5], Wei Feng[2], Yanhao Zhu[5], Yang Xiao[2], Xiaojian Xu[6], Yuji Wang[7], Daohua Jiang[3]*, Huting Wang[5]*, Zhenming Liu[1,8,9]*, and Bo Huang[5,7]*

[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, China.
[2]Beijing StoneWise Technology Co Ltd., Haidian Street #15, Beijing, 100080, China.
[3]Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China.
[4]State Key Laboratory of Green Biomanufacturing, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing, China.
[5]NeoPrimeTech Biology, Beijing, China
[6]Beijing Neurosurgical Institute, Capital Medical University, Beijing, China.
[7]College of Pharmaceutical Sciences, Capital Medical University, Beijing 100069, P. R. China.
[8]Key Laboratory of Xinjiang Endemic Phytomedicine Resources Ministry of Education; School of Pharmacy, Shihezi University, Shihezi 832003, Xinjiang, China.
[9]Beijing Key Laboratory of AI-Driven Drug Discovery and Development, Beijing, China.

[†]These authors contributed equally to this work.

*Corresponding author(s). E-mail(s): bohuang_011@163.com; zmliu@bjmu.edu.cn; wanghuting@stonewise.cn; jiangdh@iphy.ac.cn; wanghan0501@outlook.com

Contents

1. Supplementary Results

1.1. Controllable Molecular Shape Generation via NCI and Anchor Conditioning

To clarify the role of NCI and anchor specifications, we further analyze their effects on interaction recovery and spatial control. Unlike conventional metrics, these specifications are mainly designed to guide desired protein–ligand interactions and molecular placement within the binding pocket.

NCI and anchor specifications mainly contribute to binding mode fidelity and shape control, rather than to topology- or conformation-related metrics. As shown in Table1 and Extended Data Table 2, adding NCI and anchor specifications does not lead to substantial changes in conventional metrics such as Vina/Glide score, QED, SAS, or the percentage of molecules with ECFP_TS > 0.5. This observation suggests that NCI and anchor constraints do not fundamentally alter the intrinsic properties of the generated molecules, including their topological features, conformational quality, or predicted binding affinities. Instead, their primary contribution lies in improving the model's ability to satisfy specified non-covalent interactions between the generated molecules and the binding pocket, while also providing a means to control the spatial shape and placement of generated molecules. These aspects are not adequately captured by conventional molecular quality or affinity-based metrics.

NCI and anchor specifications improve % NCI in generated molecules. The main purpose of introducing NCI specifications is to guide the model to generate molecules that recapitulate desired non-covalent interactions, such as hydrogen bonds and hydrophobic contacts, with specific pocket residues. As shown in Table 1, compared with the *de novo* setting without NCI and anchor constraints, incorporating these specifications increases the % NCI from 36.0% to 43.1%. This improvement directly addresses an important practical need in structure-based drug design: generated molecules should be able to preserve critical protein–ligand interactions specified by prior knowledge, rather than relying on chance to form the desired binding mode.

NCI and anchor specifications steer the spatial geometry of generated molecules within the binding pocket. Beyond improving interaction recovery, NCI and anchor specifications provide an effective mechanism for controlling the spatial geometry of generated molecules in the binding pocket. To demonstrate this capability, we added Supplementary Figure S6, which compares molecules generated under two settings: (1) without NCI and anchor specifications, and (2) with NCI and anchor specifications derived from a reference ligand, either co-crystallized or custom-designed. Specifically, NCI specifications are extracted from the reference protein–ligand complex using the ProLIF package by identifying non-covalent interactions, while anchor atoms are defined as protein pocket atoms located within 4 Å of any reference ligand atom. The results show

that introducing NCI and anchor constraints can systematically guide the spatial placement and overall geometry of generated molecules within the binding pocket.

Importantly, the reference structure does not need to be carefully optimized or chemically sophisticated. Rather, it can serve simply as a spatial placeholder occupying the desired sub-pocket region. In practice, users may place an arbitrary scaffold or dummy ligand in the target region to define the corresponding NCI and anchor constraints, without requiring precise chemical design of the reference molecule. This feature is particularly useful for ligand design against specific sub-pockets, as it allows users to indicate a desired binding region through such a placeholder structure and subsequently use the derived constraints to guide molecular generation.

1.2. Ablation Studies of UniLingo3DMol

For the convenience of retrieving ablation experiments, we summarized the ablation experiments involved in this article, as shown in Supplementary Table S2.

1.2.1. Effectiveness of Multi-stage Training

To investigate the impact of each training stage, we independently ablated Stage 1, Stage 2, and Stage 3 on the *de novo* design scenario (based on the DUD-E dataset), with results summarized in Supplementary Table S3.

Omitting Stage 1 almost degraded all metrics. The % ECFP_TS > 0.5 decreased from 71.6% to 7.4%, the Glide M score increased from -7.0 to -5.0, the Vina M score increased from -6.7 to -4.1, and the % NCI decreased from 36.0% to 3.1%. This simultaneous decline across structural, binding, and interaction metrics is consistent with the role of Stage 1, which leverages the diverse chemical space within its dataset to provide superior initialization parameters for ligand generation.

Omitting Stage 2 primarily weakened UniLingo3DMol's ability to generate drug-like molecules. This was evidenced by a decrease in the number of drug-like molecules generated dropping from 72,421 to 47,470, a reduction of approximately 35%. It also weakened the quality of binding modes within pockets, as the Glide M score increased from -7.0 to -6.5 and the Vina M score increased from -6.7 to -6.1, and reduced the ability to reproduce known active compounds, with the % ECFP_TS > 0.5 decreasing from 71.6% to 15.7%. These declines are consistent with the properties of Stage 2 data: a large-scale dataset of pocket-ligand complexes with force-field accuracy supports both the refined in-pocket binding modes and the scale of drug-like output, so its removal weakens these facets together.

Omitting Stage 3 lowered the drug-likeness of the generated molecules. Although omitting Stage 3 improved several docking metrics, including the Glide M score, the Glide D score, the Vina M score, and IMP. These docking scores, however, are highly dependent

on the parameters of the force-field-based scoring functions used, such as Schrödinger's Glide module or Vina. Because Stage 3 utilized a dataset derived from experimentally validated pocket-ligand complexes, some of which exhibited lower Glide M scores, training on this dataset caused these scores to degrade. Nevertheless, Stage 3 increased the model's capability to generate drug-like molecules, raising the drug-like molecule percentage from 64.8% to 71.4%. Given our goal of generating more drug-like molecules, Stage 3 is therefore necessary.

Taken together, these ablations show that each stage plays a distinct and complementary role: Stage 1 establishes the broad chemical space prior that is most critical for reproducing known actives, Stage 2 supplies force-field-accurate binding modes and the scale of drug-like generation, and Stage 3 anchors generation in experimentally validated pocket-ligand complexes, improving structural realism and drug-likeness. All three stages are therefore retained.

1.2.2. Effectiveness of Multi-task Training

Ablation studies were performed on the *de novo* design scenario by including or excluding Task 1, with results presented in Supplementary Table S4.

Supplementary Table S4 shows that removing Task1 degraded the model's performance across the reported metrics: the % ECFP_TS > 0.5 decreased from 71.6% to 70.6%, the Glide M score increased from -7.0 to -6.8, and the Vina M score increased from -6.7 to -6.6. Although these changes are modest, they are consistent in direction across both binding mode quality (Glide M and Vina M scores) and the reproduction of known active compounds (% ECFP_TS > 0.5), consistent with an incremental but useful contribution from the auxiliary objective.

Mechanistically, the benefit of Task 1 stems from its objective: training UniLingo3DMol to predict NCI sites and anchor sites from a given protein pocket. This capability enables UniLingo3DMol to identify key pocket residues, such as amino acids and skeletal residues in active sites. Crucially, this residue identification directs UniLingo3DMol to generate ligands that preferentially occupy target-binding sites during molecular design.

1.2.3. Comparing the Pre-training Approaches between UniLingo3DMol and Other Methods

To further clarify the relationship between our Stage 1 training method and prior denoising-based molecular pre-training approaches, we provide an additional comparison with representative works, including Uni-Mol[1] and EPT[2]. These methods are included as representative examples of coordinate-based and graph-based molecular representation learning frameworks.

Method	Uni-Mol	EPT	UniLingo3DMol
Molecular representation	Atom token with 3D continuous coordinates	Graph	DSMILES with 3D discrete coordinates
Pre-training objective	Coordinate denoising	Graph-level denoising	Autoregressive reconstruction
Task paradigm	Discriminative learning	Discriminative learning	Generative modeling

As summarized in the above table, existing approaches primarily focus on either continuous coordinate reconstruction (e.g., Uni-Mol) or graph-level discriminative denoising tasks (e.g., EPT), and are typically formulated as regression or discriminative learning problems. In contrast, UniLingo3DMol operates on a discrete DSMILES representation in which the 3D coordinates are likewise discretized, and formulates pre-training as joint autoregressive sequence reconstruction and coordinate prediction, thereby enabling a generative modeling paradigm over molecular structures.

1.3. Screening Pipelines for CBL-B Compound Design

1.3.1. Screening Pipeline for Novel Scaffold Discovery

The primary aim of the initial screening phase was the identification of novel scaffolds. Using the proposed UniLingo3DMol model, we generated molecules based on two retained fragments (Regions A and C; Figure 2c). In this round, the UniLingo3DMol model was used to sample 3 million times. After filtering out outputs that violated the DSMILES syntax as well as redundant SMILES entries, we obtained 697,204 unique molecules. This resulting set comprises a diverse CBL-B compound library, particularly exhibiting high variability in the B and D regions.

We performed virtual screening of the compound library using a customized pipeline (Figure 3b). First, molecules were filtered based on key chemical properties: QED > 0.3, SAS < 5, and number of rotatable bonds < 10. This initial filtering yielded a refined library of 186,434 unique molecules. Second, these compounds underwent molecular Glide min-in-place docking into the binding site of the target protein (PDB: 8GCY) using Schrödinger. Subsequently, conformational filtering was applied using the TED toolkit[3] to retain molecules with a torsional energy score below 1, resulting in a library of 12,143 unique molecules. Third, a binding-mode-based filter was implemented to further reduce the library. Specifically, ProLIF was employed to retain only molecules forming hydrogen bond with residue F263 of the pocket, which yielded a final library of 3,356 molecules.

Subsequently, we computed the CSK scaffolds for all molecules in the final compound library. To prioritize novelty, we filtered out molecules that shared the same CSK scaffold

as the co-crystallized ligand of protein 8GCY. The remaining molecules were then clustered based on their CSK scaffolds, meaning that molecules with identical scaffolds were assigned to the same cluster. This process resulted in 344 clusters. Within each cluster, molecules were ranked in ascending order according to their Glide M score. Finally, the top-ranked molecule from each cluster was selected as its representative. Additionally, the visualizations for the entire set of representative molecules are provided in Supplementary Data (Zenodo repository https://zenodo.org/records/20265793).

1.3.2. Screening Pipeline for Lead Compound Optimization

In contrast to the initial screening round, the second round employed the UniLingo3DMol model for the generation of molecules, specifically retaining Regions A, B, and D (Figure 2d). In this round, we also employed the UniLingo3DMol model to sample 3 million times. Due to the use of a larger retained fragment compared to the first round, the number of generated molecules with identical SMILES increased. After filtering out outputs that violated the DSMILES syntax and removing redundant SMILES entries, we obtained 467,920 unique molecules.

A virtual screening workflow (Figure 4a), analogous to that used in the first round, was subsequently applied, incorporating three distinct filter types: chemical property, molecular conformation, and binding mode. After the chemical property filter, 263,574 molecules were retained, while the molecular conformation filter yielded 64,885 molecules. Notably, the binding-mode-based filter criteria were distinct in this round. Molecules were retained exclusively if they formed a hydrogen bond interaction with F263 while simultaneously establishing a salt bridge with E268. This final filter resulted in a library of 1,331 molecules.

Finally, we computed the CSK scaffold for each molecule in this library and clustered molecules sharing identical scaffolds, resulting in 91 clusters. Molecules within each cluster were ranked in ascending order based on their Glide M score. The top-ranked molecule from each cluster was selected as its representative. Visualizations of all representative molecules are provided in Supplementary Data (Zenodo repository https://zenodo.org/records/20265793).

2. Supplementary Methods

2.1. UniLingo3DMol Development

2.1.1. Training algorithms for UniLingo3DMol

UniLingo3DMol implemented a hierarchical three-stage training schema that includes the training Stage 1, Stage 2, and Stage 3.

In Stage 1 (Supplementary Algorithm S1), UniLingo3DMol underwent training in ligand reconstruction using perturbed ligand data $D_l^{\dagger}$ and DSMILES-formatted ligand data D_l together with their coordinates. Stage 1 was conducted for 500 epochs, with UniLingo3DMol autoregressively predicting three key components through dedicated generation heads: ligand token head, ligand pointer head, and ligand position head. The composite loss $\mathcal{L}_{\mathrm{PT}}$ combined the token prediction loss ($\mathcal{L}_{\mathrm{token}}$), pointer loss ($\mathcal{L}_{\mathrm{ptr}}$), and positional loss ($\mathcal{L}_{\mathrm{pos}}$), with parameters updated via AdamW optimization after each epoch.

In Stage 2 (Supplementary Algorithm S2), UniLingo3DMol was initialized with the parameters from Stage 1 to leverage transfer learning benefits, a common practice in deep learning. UniLingo3DMol interleaved three distinct tasks over 200 epochs on the pocket-ligand complex data generated through force field. The first task (Task 1) predicted NCI sites and anchor sites from pockets, the second task (Task 2) conducted unconstrained ligand generation, and the third task (Task 3) performed constrained ligand generation with NCI/anchor awareness. Each epoch iterated through all three tasks sequentially, calculating task-specific losses ($\mathcal{L}_{\tau}$). Crucially, parameter updates occurred after each individual task execution rather than per epoch, enabling focused gradient propagation for each objective. Both stages employed data shuffling and AdamW optimization, with the multi-task phase demonstrating a curriculum learning strategy through its alternating constraint-enabled and constraint-free generation tasks.

In Stage 3 (Supplementary Algorithm S2), UniLingo3DMol was initialized with the parameters from Stage 2 and fine-tuned on the target dataset, which consisted of protein-ligand complex data with experimental accuracy, using the same three tasks as in Stage 2. To prevent catastrophic forgetting and improve UniLingo3DMol's generalization ability, Stage 3 was only carried out for 10 epochs.

The definitions of abovementioned losses are as follows:

- **Token Loss** ($\mathcal{L}_{\mathrm{token}}$): It measures the discrepancy between the predicted and ground-truth DSMILES-formatted molecular tokens.
- **Pointer Loss** ($\mathcal{L}_{\mathrm{ptr}}$): It measures the discrepancy between the predicted and ground-truth tokens' connection relationships,
- **Coordinate Loss** ($\mathcal{L}_{\mathrm{pos}}$): It consists of global and local coordinates losses:

- $\mathcal{L}_r^{\text{local}}$: It measures the error between the predicted and ground-truth radial distances.
- $\mathcal{L}_\theta^{\text{local}}$: It measures the discrepancy between the predicted and ground-truth bond angles.
- $\mathcal{L}_\phi^{\text{local}}$: It evaluates the difference between the predicted and ground-truth dihedral angles.
- $\mathcal{L}_{\text{pos}}^{\text{global}}$: It evaluates the difference between the predicted and ground-truth atomic global coordinates.

- **NCI Loss** ($\mathcal{L}_{\text{NCI}}$): It measures the discrepancy between the predicted and ground-truth NCI sites.
- **Anchor Loss** ($\mathcal{L}_{\text{Anchor}}$): It measures the discrepancy between the predicted and ground-truth anchor sites.

Here, $\mathcal{L}_{\text{NCI}}$ and $\mathcal{L}_{\text{Anchor}}$ were calculated using binary cross-entropy loss functions, and all other losses were calculated using the cross-entropy loss functions.

2.1.2. Hyperparameters of UniLingo3DMol

The three-stage UniLingo3DMol training experiment was performed on a distributed computing cluster with Nvidia H20s (Supplementary Table S9). All stages consistently employed the AdamW optimizer with identical learning rates and weight decay regularization.

2.1.3. Sequence Augmentation Algorithms

The Supplementary Algorithm S3 transforms an original fragment sequence FS_{raw} into a randomized yet structurally valid sequence FS_{trans} by iteratively selecting and appending fragments while adhering to connection constraints. Initially, the length n of FS_{raw} was determined, and FS_{trans} was initialized as an empty list. In each iteration, if FS_{trans} is empty, a fragment f is randomly chosen from FS_{raw}, added to FS_{trans}, and removed from FS_{raw}. For subsequent steps, the algorithm identifies connection constraints satisfied by the current FS_{trans}, randomly selects a compatible fragment f from FS_{raw}, transforms f into FS_{trans} using an asterisk constraint, appends f_{trans} to FS_{trans}, and removes f from FS_{raw}, decrementing n until all fragments are processed. Finally, FS_{trans} undergoes a legality check to ensure compliance with structural rules, ensuring the output is both stochastically augmented and valid.

The Supplementary Algorithm S4 transforms an original fragment sequence (FS_{raw}) into a new sequence (FS_{trans}) by retaining a specified number of fragments (n_r) while modifying others. First, it checks if retention is unnecessary ($n_r \geq FS_{\text{raw}}$) and returns an empty sequence. Otherwise, it initializes FS_{trans}, randomly selects a starting index, and iteratively processes fragments in two phases:

- **Preservation phase**: Fragments are randomly selected based on unsatisfied connection constraints and added to FS_{trans} until n_r fragments are retained.
- **Conversion phase**: Remaining fragments are modified using an asterisk constraint based on satisfied connection requirements. At each step, the selected fragment is removed from FS_{raw}, and the counters are updated. The algorithm ends when all fragments are processed, followed by a legality check to validate the structural consistency of FS_{trans}. Randomized selection and constraint-driven retention ensure diversity while preserving critical sequence relationships.

2.2. Molecular Retrieval-Augmented Generation using UniLingo3DMol

We enhance the generative pipeline by incorporating molecular fragment retrieval, which allows the model to access a library of 3D interaction pairs between amino acids and molecular fragments (Supplementary Figure S5). Each entry in this library includes the 3D structures of an amino acid and a corresponding molecular fragment, annotated with quantum mechanically accurate interaction energies computed using the machine-learned force field AIMNet2[4], as well as intuitive interaction types defined by geometric criteria via ProLIF[5]. All interaction pairs are derived from known protein-ligand complexes (Level 0 from RComplex database), ensuring that the generation process is grounded in chemically and biologically relevant knowledge, thereby minimizing the risk of invalid or unrealistic molecular designs.

During retrieval, the library is queried using protein residue indices and desired interaction types, optionally focusing on backbone interactions where all types of amino acids will be screened, to identify fragments whose amino acid partners have conformations resembling the query residue and are compatible with the protein pocket. Candidate fragments are selected based on energy and geometry criteria and then retained as fragments for molecule generation.

2.2.1. Retrieval Library Construction

The retrieval library is constructed by using RComplex Level 0 data (Methods Section 4.7). For each complex, the binding pocket is defined as all residues with at least one atom within 6 Å of the ligand. Ligands are decomposed into fragments by cleaving non-cyclic, non-conjugated, or non-terminal bonds. Any fragment consisting of only one heavy atom is merged into its most adjacent fragment, and hydrogens are added at the cleavage points.

For each amino acid residue in the binding pocket, RDKit[6] is used to add hydrogens to the backbone nitrogen (N) and alpha carbon (Cα) to neutralize unwanted charges. ProLIF[5] is then employed to detect interactions between the amino acid residue and each ligand fragment. For each amino acid residue–ligand fragment pair, energy minimization is performed using the GAFF2[7] force field and OpenMM while constraining the amino acid

residue conformation. The 3D conformations of the amino acid residue and ligand fragment are recorded, along with the following features: the interaction fingerprints (IFP) from ProLIF; the RMSD of ligand fragment heavy atoms before and after minimization (min_{rmsd}); and the non-covalent interaction energies (NCIE) computed with the AimNet2 force field, both before ($NCIE_{ori}$) and after minimization ($NCIE_{min}$).

After extracting all interaction pairs, they are grouped by amino acid type. For each type, an arbitrary entry is selected as a reference, and the remaining entries are aligned to it with respect to the backbone atoms (N, Cα, C) as the reference. The resulting transformation is applied to the corresponding fragment to preserve complex geometry. A distance matrix based on side-chain heavy atoms is used to cluster the amino acid residues into conformers, and each conformer's ID and centroid status are recorded.

The final library entry includes: the amino acid residue structure in PDB format; the original and minimized fragment structures in SDF format; the amino acid conformer ID; an indicator of whether the amino acid is the centroid of its cluster; RMSD of ligand fragment heavy atoms between conformation before and after minimization; NCIE before and after minimization; IFPs between the amino acid residue and ligand fragment in the original conformation; and an indicator of whether the IFP involves backbone atoms.

2.2.2. Retrieval Method

The retrieval process is query-based, using an IFP site defined by a protein residue index (chain ID and residue ID) and the desired IFP type between the target residue and the retrieved fragment. Users may optionally specify that the IFP should involve the residue's backbone; in such cases, the amino acid type is ignored, and data from all residue types are considered. The method compares the query residue conformation to those in the database to identify entries with similar amino acids conformers, while ensuring that the retrieved fragment does not clash with the protein pocket via a docking score.

The retrieval proceeds as follows. First, the dataset corresponding to the specified residue type is selected; if backbone IFPs are targeted, the dataset includes all residue types and is filtered for backbone interactions. The query residue conformation is compared to all cluster centroids by calculating the RMSD of matching heavy atoms after backbone alignment. Conformer clusters with the closest-matching centroids are selected. From the selected entries, those with $NCIE_{ori} < 0$ are retained, as are those with $NCIE_{ori} \geqslant 0$ that decrease to $NCIE_{min} < 0$ after minimal adjustment (constrained minimization RMSD < 2 Å). Further filtering is applied based on the target IFP type if specified. The candidate amino acid is aligned to the query using backbone atoms, and the fitting RMSD is recorded with a precision of 0.05 Å. For full-residue matching, alignment is refined using all matching heavy atoms; for backbone-only cases, only backbone heavy atoms are used. The resulting transformation is applied to the fragment.

Potential growth sites are identified and adjusted. These sites are initially defined at original bond cleavage points. For each site, a pseudo-atom is placed 1.5 Å along the vector from the heavy atom to its connected hydrogen. The minimum distance from this pseudo-atom to all heavy atoms in the protein pocket and ligand (excluding the site itself and neighbors) is computed. If no original cleavage sites exist, other carbon atoms with attached hydrogens are considered as potential sites. Sites with a minimum distance $\geqslant$ 2.5 Å are retained. If more than three qualify, the top three are selected based on the largest minimum distance for original cuts, or to maximize spatial diversity for new sites. If no site qualifies, the primary site is used. The maximum of the minimum distances across selected sites (max_min_space) is recorded and rounded to 0.1 Å. Docking scores are computed using Smina[8] (Vinardo scoring) to ensure no steric clashes with the protein pocket. Negative scores are rounded to the nearest multiple of 2 kcal/mol and adjusted by subtracting 0.001 to avoid zero values. The positioning direction angle is calculated as the angle between the vector from the query residue's center of mass (COM) to the fragment's COM and a reference vector from the residue COM to the pocket COM (or a reference ligand COM if provided). This angle is rounded to the nearest even degree.

Growth direction angles are computed for each selected growth vector (from the heavy atom to its connected hydrogen at the growth site). The angle between this vector and a reference growth vector is calculated, and the smallest angle is selected as the best growth angle (rounded to the nearest even degree). The reference vector depends on the fragment's distance to the reference COM: if the distance is less than 2 Å, the reference vector points from the reference COM to the growth site; otherwise, it points from the growth site to the reference COM.

Candidates are filtered to exclude those with position angles > 120°, best growth angles > 120°, docking scores $\geqslant$ 1.0, more than 30 heavy atoms, or no valid growth sites. Sulfur-containing fragments may be optionally excluded to avoid excessive amount of sulfonyl group. Duplicates are removed based on canonical SMILES (ignoring isotopes). The remaining fragments are ranked by docking score (ascending), fitting RMSD (ascending), position angle (ascending), number of IFPs (descending), and reference NCIE (ascending, using $NCIE_{min}$ if adjusted).

2.2.3. RAG with Molecular Generation

To evaluate the RAG strategy, we applied it to the DUD-E dataset (Supplementary Table S7). For each target, ProLIF was used to identify key IFP patterns (hydrogen bonds, π-stacking, and salt bridges) in the cocrystal structure. Protein residues involved in these interactions were treated as query sites. For each site, fragments were retrieved using the method described above. Due to resource constraints, a single fragment was selected as the

initial seed for molecular generation. This selection involved an additional ranking across all retrieved fragments based on docking score and the total number of IFPs formed with the protein pocket, with the top-ranked fragment chosen for subsequent generation.

2.3. Model Evaluation

2.3.1. Evaluation Sets for *De Novo* Design

For the comprehensive evaluation of *de novo* molecular design strategies, the DUD-E dataset served as the foundational resource. This dataset provided 102 unique protein targets and 22,886 active compounds whose activities were previously reported in the literature. A refinement during data preparation involved the substitution of one protein structure: PDB ID 2H7L was found to be obsolete in the PDB, and as per the guidance of the RCSB PDB, the structure 4TRJ was consequently utilized.

2.3.2. Evaluation Sets for Fragment-retained Design

Three evaluation sets were constructed from the DUD-E dataset to assess model performance in the fragment-retained scenario. The core methodology involved identifying high-frequency fragments derived from co-crystallized ligands and subsequently classifying active molecules based on their presence. Specifically, both co-crystallized ligands (derived from DUD-E crystal structures) and known active molecules were systematically fragmented by breaking the non-ring single bonds. This process generated a set of molecular fragments for each molecule, and we subsequently conducted a frequency analysis of these fragments. For each target, we identified the fragments found in the co-crystallized ligands that also appeared in the active compounds set. Fragments that appeared more than five times in the active molecule set for a given target were labeled as conserved fragments. This threshold was empirically chosen to ensure the selection of well-represented motifs.

According to the presence or absence of these conserved fragments, three fragment-retained design evaluation sets were obtained:

- **Single-fragment-retained evaluation set**: 144 samples from 67 DUD-E proteins. Each sample contains one identified conserved fragment. For each sample, the active molecule corresponding to the conserved fragment is used as the set of known active compounds that the model needs to reproduce.
- **Two-fragment-retained evaluation set**: 79 samples from 30 DUD-E proteins. Each sample consists of two identified conserved fragments. Since the DUD-E active compounds set can hardly contain two identified conserved fragments simultaneously, we did not evaluate the active compounds reproducibility performance of models.

- **Three-fragment-retained evaluation set**: 19 samples from 9 DUD-E proteins. Each sample contains three identified conserved fragments. Following the same rationale applied to the two-fragment-retained set, we did not evaluate the reproducibility of the active compounds.

2.3.3. Definition of Evaluation Metrics

The metrics used in this study are as follows:

- **# Generated molecules**: The sum of total number of generated molecules across 102 targets, with an upper limit of 1000 per target.
- **MW**: The average molecular weight of all generated molecules.
- **QED**: The averaged quantitative estimate of drug-likeness of all generated molecules.
- **SAS**: The averaged synthetic accessibility score of all generated molecules.
- **RAS**: The average retrosynthetic accessibility score of all generated molecules.
- **% drug-like molecules**: The percentage of molecules with QED > 0.3, SAS < 5, and RAS>0.5 over all generated molecules.
- **BL JSD**: The average statistical dissimilarity between the generated molecules and the CSD[9] reference for 14 common bond length types.

$$\text{Bond length JSD} = \frac{\sum_i \frac{1}{2} D_{KL}(P\|M) + \frac{1}{2} D_{KL}(Q\|M)}{14}$$

  where i represents one of the 14 common types of bond lengths, including C-C, C=C, C:C, C#C, C-N, C=N, C#N, C:N, C-O, C=O, C:O , C-S, C=S and C:S. The symbol P and Q denote the bond length values from generated molecules and the CSD reference, respectively. The term M refers to the intermediate distribution, defined as $M = \frac{1}{2}(P + Q)$.
- **BA JSD**: The average statistical dissimilarity between the generated molecules and the CSD[9] reference for 11 common bond angle types.

$$\text{Bond angle JSD} = \frac{\sum_i \frac{1}{2} D_{KL}(P\|M) + \frac{1}{2} D_{KL}(Q\|M)}{11}$$

  where i represents one of the 11 common types of bond angles, including CC=C, CC=O, CCC, CCO, CN=C, CNC, COC, CSC, NC=O, NCC, and OC=O. The symbol P and Q denote the bond angles values from generated molecules and the CSD reference, respectively. The term M refers to the intermediate distribution, defined as $M = \frac{1}{2}(P + Q)$.
- **DA JSD**: The average statistical dissimilarity between the generated molecules and the CSD[9] reference for 6 common dihedral angle types.

$$\text{Dihedral angle JSD} = \frac{\sum_i \frac{1}{2} D_{KL}(P\|M) + \frac{1}{2} D_{KL}(Q\|M)}{6}$$

where i represents one of the 6 common types of dihedral angles, including CCCC, cccc, CCCO, OCCO, Cccc, and CC=CC. The symbol P and Q denote the dihedral angles values from generated molecules and the CSD reference, respectively. The term M refers to the intermediate distribution, defined as $M = \frac{1}{2}(P + Q)$.

- **Vina M score**: The average of Vina score using "optimize" method of AutoDock. During calculation, we only considered drug-like molecules with a negative score.
- **Vina D score**: The average of Vina score using "docking" method of AutoDock. During calculation, we only considered drug-like molecules with a negative score.
- **Glide M score**: The average of Glide score using "min-in-place" method of Schrödinger Glide module. During calculation, we only considered drug-like molecules with a negative score.
- **Glide D score**: The average of Glide score using "docking" method of Schrödinger Glide module. During calculation, we only considered drug-like molecules with a negative score.
- **% IMP**: The percentage of generated drug-like molecules whose Glide M score is lower than their respective co-crystal ligand for the respective target. During calculation, we only considered drug-like molecules with a negative Glide M score.
- **NCI recovery rate**: This metric is a measurement of the fraction of non-covalent interactions in the crystal ligand pose that are successfully replicated in the generated molecule pose using ProLIF, defined as:

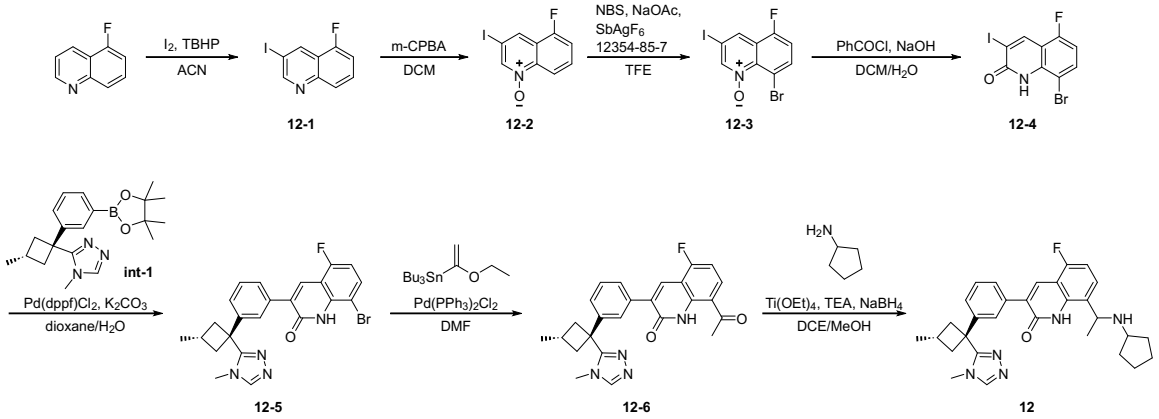
$$\text{NCI Recovery} = \frac{\sum_{i,r} \min\left(C_{i,r}, P_{i,r}\right)}{\sum_{i,r} C_{i,r}}$$

where $C_{i,r}$ is the count of type-i interaction that crystal ligand formed with residue r, and $P_{i,r}$ is the count of type-i interaction that a generated molecule formed with residue r. Here, we considered interactions including hydrogen bond acceptor/donor, halogen bond acceptor/donor, π-π stacking, cation-π, π-cation, anionic and cationic interactions were calculated by ProLIF.

- **% ECFP_TS > 0.5**: The ratio for which at least one generated drug-like molecule has a Tanimoto similarity greater than 0.5 (based on ECFP4[10] fingerprints) to any known active compound for that target.

2.4. Compound Synthesis

2.4.1. Synthesis of Example 12



2.4.2. Synthesis of Intermediate 12-1

5-Fluoroquinoline (10 g, 68.03 mmol), iodine (20.7 g, 81.50 mmol), and 70 wt% *tert*-butyl hydroperoxide aqueous solution (43.7 g, 339.89 mmol) were dissolved in acetonitrile (100 mL). The reaction mixture was stirred at 80°C for 16 hours. After cooling to room temperature, the mixture was diluted with saturated *aq.* Na_2CO_3 until pH was adjusted to 9, and extracted with ethyl acetate (200 mL × 3). The combined organics washed with brine (30 mL × 3), dried over Na_2SO_4 and concentrated in vacuo to give crude product. The crude product was purified by silica gel column chromatography (eluent: PE/EA = 8:1) to give **12-1** (10.7 g, 57.6% yield) as a yellow solid, MS m/z (ESI): 274 [M+1].

2.4.3. Synthesis of Intermediate 12-2

To a mixture of **12-1** (10.7 g, 39.19 mmol) in dichloromethane (200 mL) was added *m*-chloroperbenzoic acid (13.56 g, 78.38 mmol) in batches under stirring at 0°C. The reaction mixture was stirred at room temperature for 16 hours. The mixture was diluted with H_2O (100 mL) and adjusted pH to 9~10 by *aq.* NaOH (1 N) at 0°C. The mixture was extracted with dichloromethane (150 mL × 3), and the organic phases were combined, dried over Na_2SO_4, evaporated and the residue was purified by silica gel column chromatography (eluent: PE/EA = 2:1) to afford **12-2** (5.0 g, 44.1% yield) as a yellow solid, MS m/z (ESI): 290[M+1].

2.4.4. Synthesis of Intermediate 12-3

To a mixture of **12-2** (5.0 g, 17.30 mmol), dichloro(5-methylcyclopentadienyl)rhodium(III) dimer (536 mg, 0.87 mmol), silver hexafluoroantimonate (1.19 g, 3.47 mmol) and sodium acetate (284 mg, 3.46 mmol) in trifluoroethanol (90 mL) was added *N*-bromosuccinimide (4.0 g, 22.47 mmol) at room temperature. The reaction mixture was stirred at 50°C for 16 hours. The reaction mixture was concentrated, and the residue was purified by silica gel

column chromatography (eluent: PE/EA = 4:1) to afford **12-3** (1.4 g, 21.9% yield) as a yellow solid, MS m/z (ESI): 368 [M+1].

2.4.5. Synthesis of Intermediate 12-4

12-3 (1.4 g, 3.80 mmol) was dissolved in a 30 mL mixture of dichloromethane and water (V/V = 1:1). Sodium hydroxide (304 mg, 7.60 mmol) was added, followed by the dropwise addition of benzoyl chloride (643 mg, 4.56 mmol) under stirring at 0°C. The reaction mixture was stirred at room temperature for 1 hour. After the reaction was completed, the precipitated solid was collected by filtration, and the filter cake was washed with water (5 mL × 3), dried in vacuo to afford **12-4** (750 mg) as an off-white solid, MS m/z (ESI): 368 [M+1].

2.4.6. Synthesis of Intermediate 12-5

A mixture of **12-4** (600 mg, 1.63 mmol), **int-1** (692 mg, 1.96 mmol), [1,1'-bis(diphenylphosphino)ferrocene] palladium dichloride (119 mg, 0.16 mmol) and potassium carbonate (450 mg, 3.26 mmol) in 12 mL of dioxane and water (V/V = 5:1) was stirred at 80 °C under nitrogen for 2 hours. After cooling to room temperature, the reaction mixture was concentrated and the residue was purified by silica gel column chromatography (eluent: DCM/MeOH = 12:1) to afford **12-5** (560 mg, 73.5% yield) as a brown solid, MS m/z (ESI): 467 [M + 1].

2.4.7. Synthesis of Intermediate 12-6

A mixture of **12-5** (300 mg, 0.64 mmol), tributyl(1-ethoxyvinyl)tin (462 mg, 1.28 mmol), and bis(triphenylphosphine)palladium dichloride (90 mg, 0.13 mmol) in *N,N*-dimethylformamide (5 mL) was stirred at 85 °C under nitrogen atmosphere for 2 hours. The reaction mixture was cooled to room temperature and 2N hydrochloric acid solution (10 mL) was added. The mixture was stirred at room temperature for 1 hour. The mixture was adjusted to pH = 9 by saturated *aq.* Na_2CO_3 (10 mL × 3), extracted by ethyl acetate (20 mL × 3). The combined organics was washed by brine (10 mL × 3), dried over Na_2SO_4, evaporated and the residue was purified by silica gel column chromatography (eluent: DCM/MeOH = 10:1) to afford **12-6** (170 mg, 61.5% yield) as a yellow solid, MS m/z (ESI): 431 [M+1].

2.4.8. Synthesis of Compound 12

12-6 (140 mg, 0.33 mmol), cyclopentylamine (56 mg, 0.66 mmol), tetraethyl titanate (226 mg, 0.99 mmol) and triethylamine (133 mg, 1.32 mmol) were dissolved in 1,2-dichloroethane (4 mL). The reaction mixture was stirred at 80°C for 16 hours. After cooling to room temperature, methanol (4 mL) was added to the reaction mixture, followed by the addition of sodium borohydride (63 mg, 1.66 mmol) in portions while stirring at room

temperature. After the addition was completed, the reaction mixture was stirred at room temperature for another 3 hours. The mixture was diluted with H_2O (20 mL) and extracted with dichloromethane (10 mL × 3). The organic phases were combined, dried over Na_2SO_4, evaporated and the residue was purified by silica gel column chromatography (eluent: DCM/MeOH = 9:1) to afford crude product. The crude product was purified by *Prep*-HPLC (column: Xbridge-C18 150 x 19 mm; eluent: ACN-H_2O (0.05%FA); gradient: 15-40%) to afford **Cmpd. 12** (25 mg, 15.4% yield) as a white solid.

Cmpd. 12 (25 mg) was separated by SFC (column: CHIRALPAK AD-H, 250 mm × 20 mm, eluent: 40% IPA (NH_4OH 0.2%), flow velocity: 40mL/min, RT1: 4.80 min, RT2: 5.95 min, column temperature: 40°C) to afford **Cmpd. 19** (8.29 mg) as a white solid. MS m/z (ESI): 500[M+1], ^{1}H NMR (400 MHz, DMSO-d_6) δ ppm: 8.28 (s, 1H), 8.02 (s, 1H), 7.76 (s, 1H), 7.61 (d, J = 7.7 Hz, 1H), 7.48 - 7.29 (m, 3H), 6.98 (dd, J = 9.8, 8.3 Hz, 1H), 4.27 - 4.18 (m, 1H), 3.22 (s, 3H), 2.92 - 2.81 (m, 3H), 2.54 (d, J = 6.6 Hz, 3H), 1.83 - 1.73 (m, 1H), 1.72 - 1.53 (m, 3H), 1.50 - 1.26 (m, 7H), 1.09 (d, J = 4.9 Hz, 3H); and **Cmpd. 20** (7.47 mg) as a white solid. MS m/z (ESI): 500[M+1], ^{1}H NMR (400 MHz, DMSO-d_6) δ ppm: 8.28 (s, 1H), 8.02 (s, 1H), 7.76 (s, 1H), 7.61 (d, J = 7.8 Hz, 1H), 7.48 - 7.29 (m, 3H), 6.98 (dd, J = 9.8, 8.3 Hz, 1H), 4.27 - 4.18 (m, 1H), 3.22 (s, 3H), 2.92 - 2.82 (m, 3H), 2.54 (d, J = 6.5 Hz, 3H), 1.83 - 1.73 (m, 1H), 1.72 - 1.53 (m, 3H), 1.50 - 1.25 (m, 7H), 1.09 (d, J = 4.8 Hz, 3H).

The compounds shown in Supplementary Table S10 were synthesized using the same methods as above.

2.5. *In-vitro* and *In-vivo* Assays

2.5.1. Expression and Purification of Human CBL-B N-domain

The gene encoding human CBL-B (UniProt code: Q13191) N-domain (Gln38-Asp427) was subcloned into a pET-28a vector fused with an 8xHis tag followed by an HRV 3C protease site at the N-terminus. The plasmids were transformed into E. coli SHuffle T7 competent cells (New England Biolabs). The cells were cultured in Luria Broth (LB) medium in a 37 °C incubator shaking at 220 rpm. When OD600 reached 0.8, the medium was supplied with 0.5 mM isopropyl β-D-thiogalactoside (Sigma). After culturing for another 22 hours at 16 °C, the cells were collected by centrifugation and were stored at -80 °C.

The cell pellets were resuspended in buffer A (20 mM Tris pH 8.0, 100 mM NaCl, 2 mM β-mercaptoethanol, and 10% glycerol) supplemented with protease inhibitors. After cell breaking by ultrasonication, the cell lysate was centrifugated for 30 min at 45000 rpm. Subsequently, the supernatant was transferred and loaded onto Ni-NTA gels in a gravity column. The resin beads were washed by 10 column volume (CV) of buffer A

supplemented with 35 mM imidazole. The protein samples were eluted by 15 mL elution buffer (20 mM Tris pH 8.0, 100 mM NaCl, 2 mM β-mercaptoethanol, and 300 mM imidazole). The His tag was removed by PreScission protease (1:50 w/w ratio) digestion for 2 h on ice. The samples were concentrated and loaded onto a Superdex 200 Increase 10/300 GL column (Cytiva) preequilibrated in buffer containing 20 mM Tris, 150 mM NaCl, 5 mM β-mercaptoethanol, pH 8.0. The peak fractions were pooled and concentrated to 20-22 mg/mL. Aliquots of the concentrated samples were flash frozen in liquid nitrogen and stored at -80 °C.

2.5.2. Crystallization, Data Collection and Structure Determination

Before crystal screening, the CBL-B N-domain samples were supplemented with 1 mM Cmpd. 7 and incubated on ice for 1 h. Initial crystals were observed in CrystalScreen Kit G6 (0.1 M HEPES pH 7.5, 10% w/v Polyethylene glycol 6,000, 5% v/v (+/-)-2-Methyl-2,4-pentanediol). This crystallization condition was optimized. Finally, 1μL protein solution was mixed with 1μL reservoir solution (0.1 M HEPES pH 7.3, 7% w/v polyethylene glycol 6,000, 5% v/v (+/-)-2-Methyl-2,4-pentanediol) for each drop, and large crystals were obtained within 7 days at 18 °C. Crystals were transferred into cryoprotectant containing mother reservoir solution supplemented with 25% ethylene glycol, then flash frozen in liquid nitrogen.

X-ray diffraction data of Cmpd. 7-CBL-B crystals were collected at beamline BL19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). X-ray diffraction data were integrated and scaled using the program HKL-2000[11]. The initial phase was obtained by molecular replacement (MR) with the CBL-B structure (PDB: 8GCY) as a reference model using the program Phaser in PHENIX[12]. Several rounds of manual model correction and refinements were performed iteratively using COOT and PHENIX refinement. Statistics for data collection and processing are shown in Supplementary Table S1.

2.5.3. CBL-B TR-FRET Assay

The biochemical activity of compounds against human CBL-B was determined using a time-resolved fluorescence resonance energy transfer (TR-FRET) assay, which utilized 30 nM SRC substrate. First, 10 μL of each test compound were transferred to a 384-well low dead volume (LDV) Echo plate. Second, compounds were diluted and transferred 150 nL to the assay plate using an Echo liquid handler, establishing 10 concentration points in duplicate. A 3.5 nM solution of human CBL-B enzyme, prepared in assay buffer, was then added at 5 μL per well to the assay plate; this was followed by centrifugation at 1500 rpm for 1 minute and a 1-minute preincubation at 25 °C. The reaction was initiated by adding 5 μL of a prepared substrate mixture to each well of the Corning 784075 assay plate, after which the plate was centrifuged again at 1500 rpm for 1 minute and incubated for 2 hours

at 25 °C. Following this incubation, 5 μL of a binding assay mix was added to each well, and the plate underwent a final centrifugation at 1500 rpm for 1 minute before being incubated for an additional 60 minutes at 25 °C. Finally, the TR-FRET signal at 520/620 nm was recorded using an Envision plate reader, and the 50% inhibitory concentration (IC_{50}) values were calculated from the recorded data using the log (inhibitor) vs. normalized response–variable slope model provided by GraphPad Prism software.

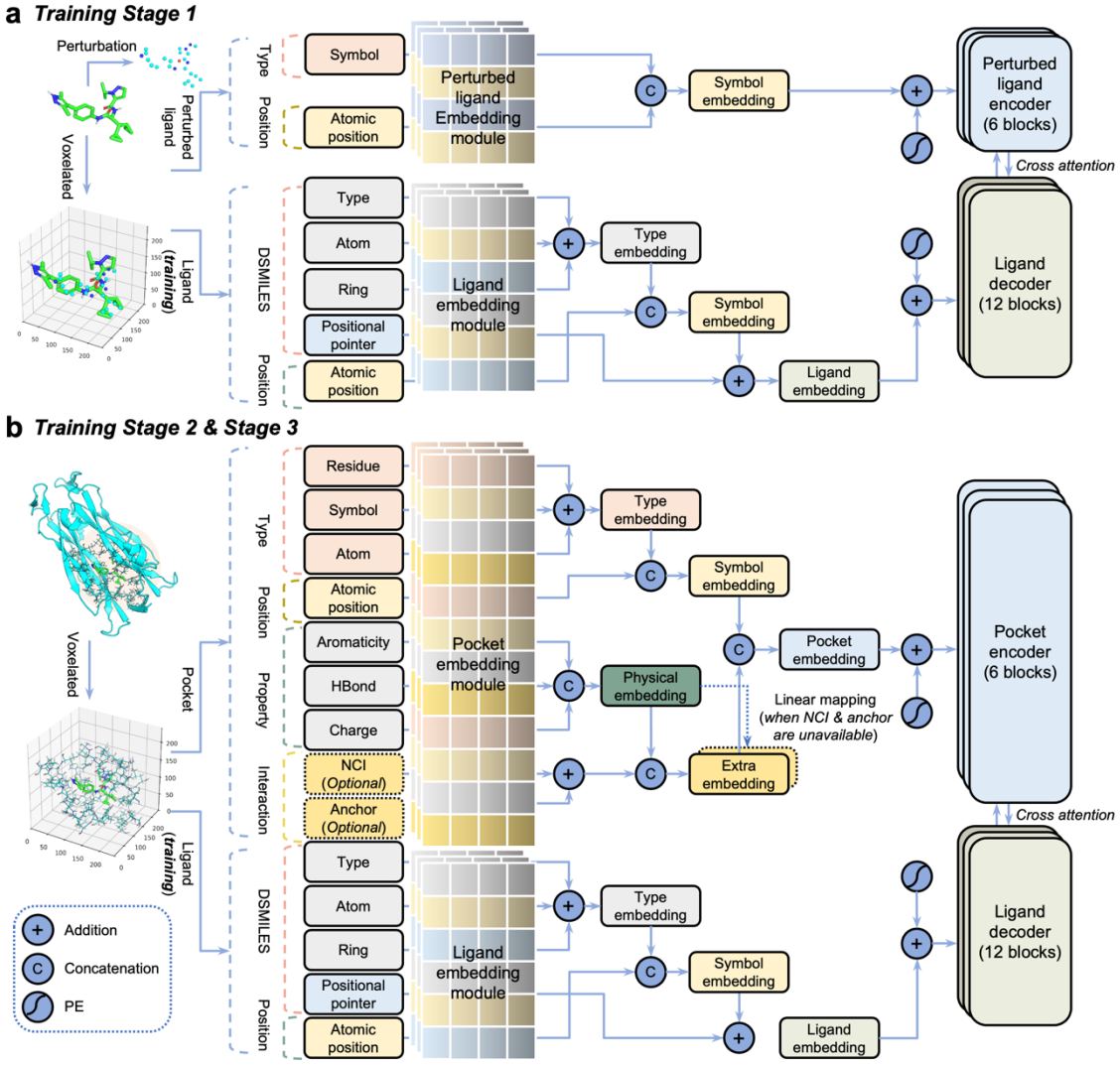
2.5.4. Jurkat IL-2 Induction Assay

Jurkat E6-1 cells were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS). For the assay, 100 μL of cell suspension (containing 1×10^5 cells) were seeded into each well of 96-well plates. The cells were then treated with varying concentrations of compounds, in triplicate, for 60 minutes. Following compound treatment, the cells were stimulated for 48 hours with 0.5 μg/mL anti-human CD3 antibody and 0.5 μg/mL anti-human CD28 antibody. Subsequently, cell culture supernatants were harvested, and the concentration of interleukin-2 (IL-2) was quantified using an enzyme-linked immunosorbent assay (ELISA) kit.

2.5.5. CT26 Murine Syngeneic Model Efficacy Studies

All animal studies were conducted at ProOnco Therapeutics Co., Ltd., under protocols approved by and in compliance with the Institutional Animal Care and Use Committee (IACUC) of the institution. CT26 tumor cells were cultured in vitro using RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS) at 37 °C with 5% CO_2. For tumor induction, BALB/c mice were subcutaneously inoculated in the right flank with 3×10^5 CT26 tumor cells suspended in 0.1 mL of phosphate-buffered saline (PBS). Drug treatments were initiated when tumor volumes reached between 50-70 mm^3. All enrolled tumor-bearing mice were randomized into study groups based on both tumor size and body weight. The investigational compound, Cmpd. 20, was administered orally as a homogeneous suspension once daily (QD), while anti-mPD-1 antibody (Clone RMP1-14, Bio X Cell) was administered intraperitoneally (i.p.) as a solution twice a week (BIW). Study animals were monitored daily for signs of morbidity and mortality. Body weight and tumor volumes were measured every 3 days. Tumor volumes were determined in two dimensions using a caliper and calculated in mm^3 using the formula: $V = (L \times W \times W) / 2$, where V represents tumor volume, L is the longest tumor dimension (length), and W is the longest tumor dimension perpendicular to L (width). Individual animals were euthanized when their tumor volume exceeded 2,000 mm^3, and the entire experiment was terminated when the mean tumor volume of the control group surpassed this threshold.

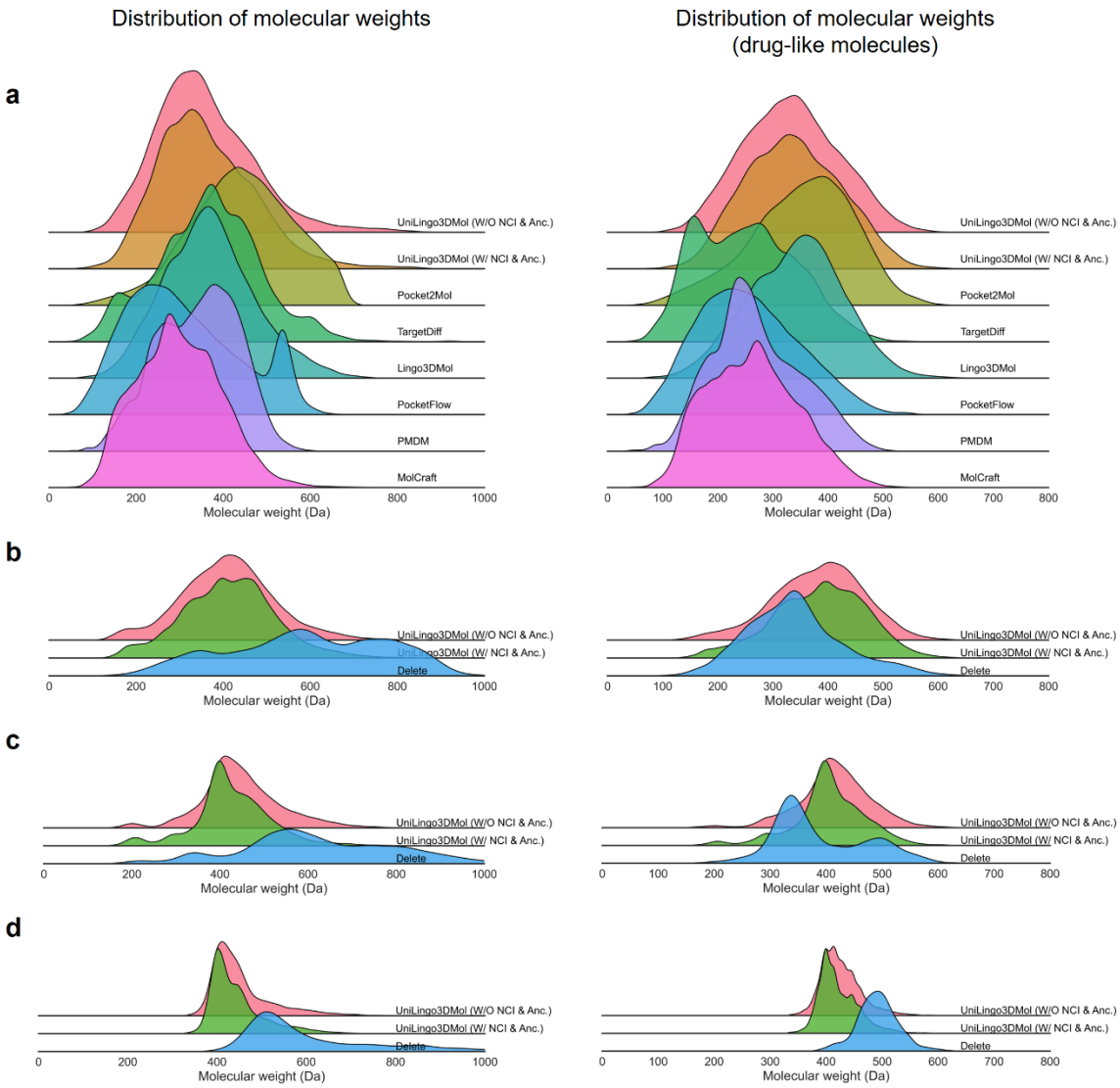
3. Supplementary Figures S1-S6



Supplementary Figure S1. Working mechanism of UniLingo3DMol across training stages.

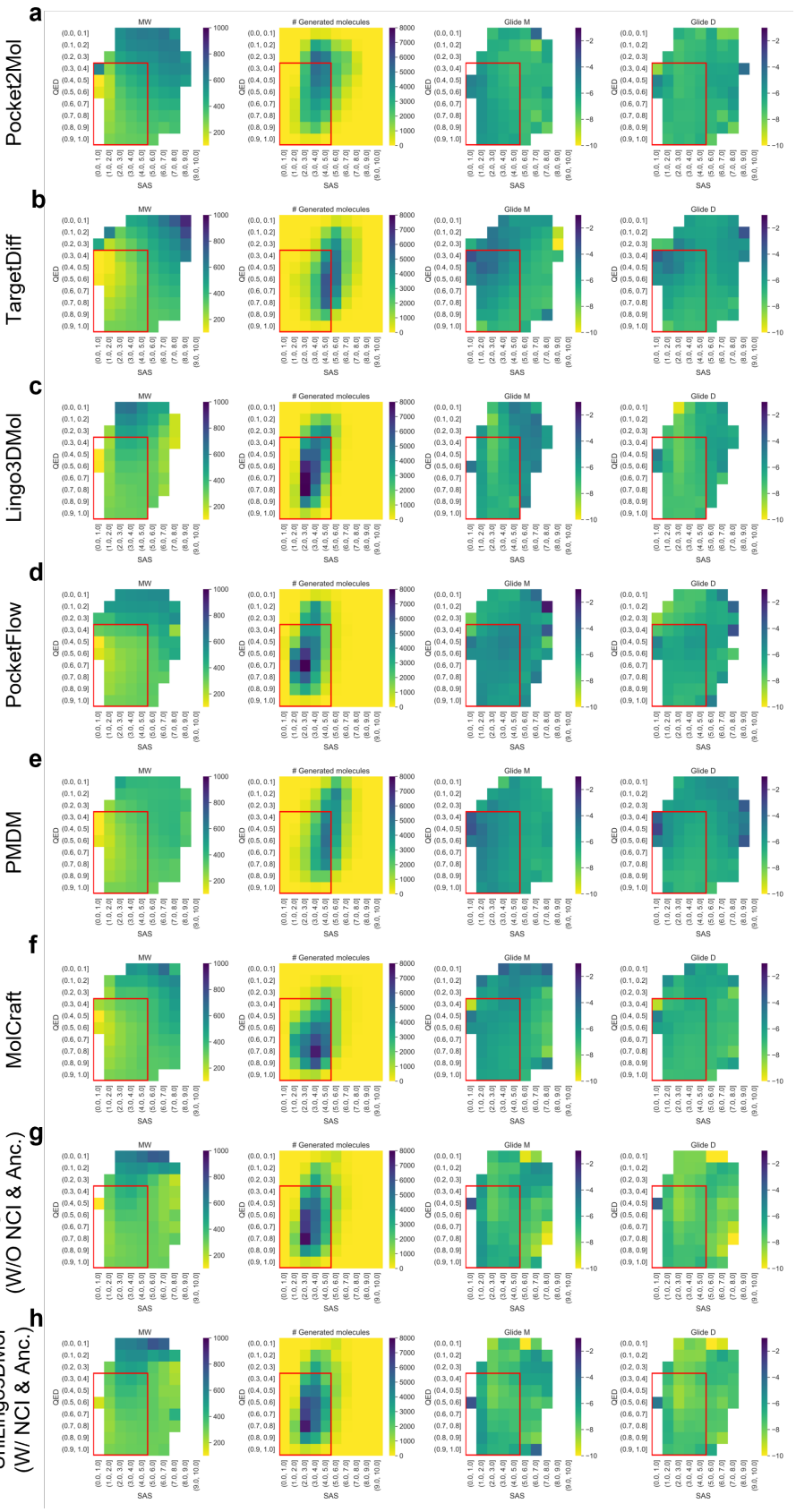
(a) Training Stage 1. The encoder processes perturbed ligand inputs, while the decoder generates ligands as DSMILES sequences with restored 3D conformations.

(b) Training Stage 2 and Stage 3. The encoder receives protein binding pockets as input.



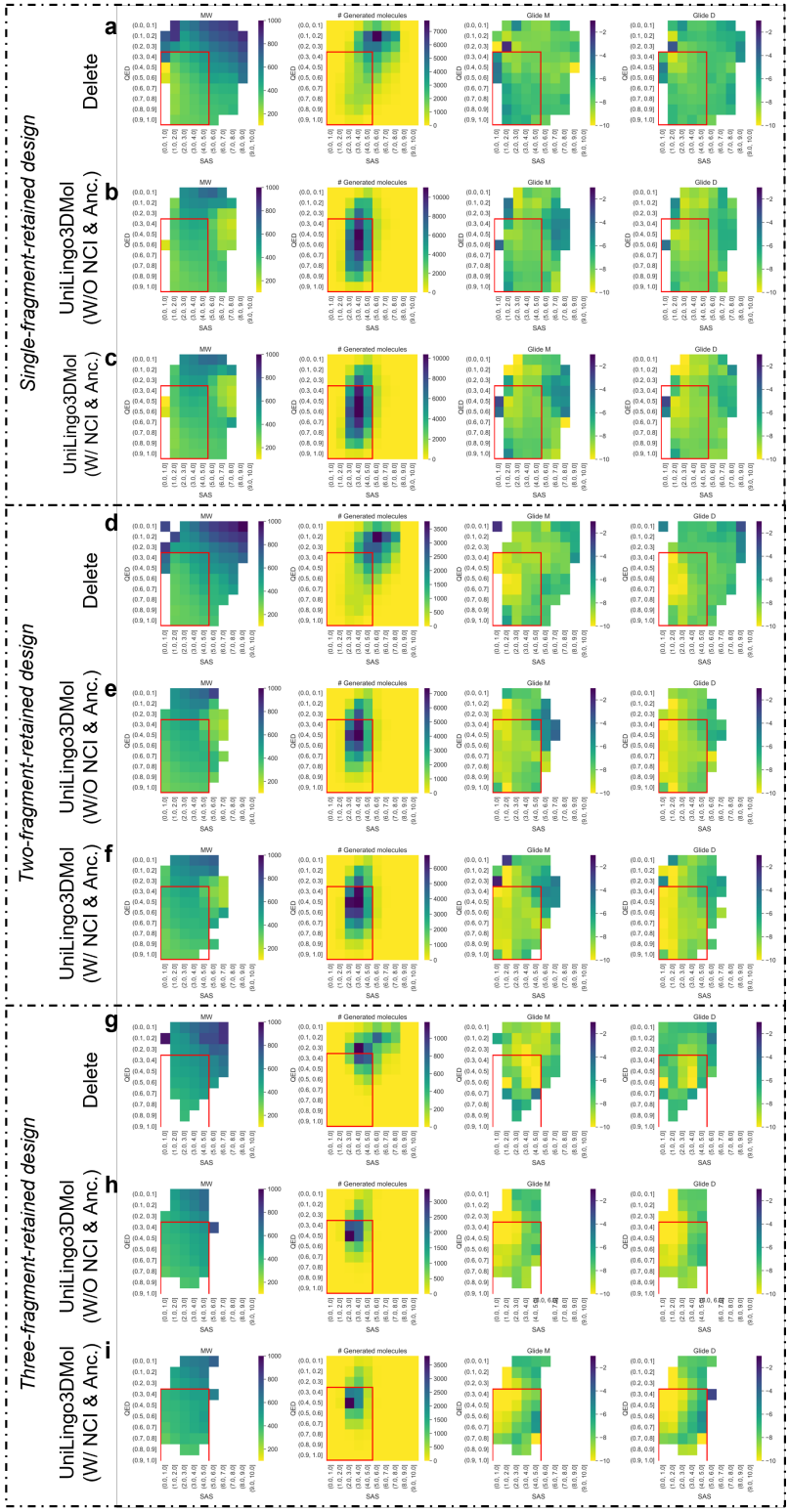
Supplementary Figure S2. Distribution of molecular weight for molecules generated by various methods on different evaluation sets.

(a-d) show results for the *de novo* design (DUD-E), single-fragment-retained, two-fragment-retained, and three-fragment-retained evaluation sets, respectively. **(a)** is supplementary to Table 1, while **(b-d)** are supplementary to Extended Data Table 2.


a
Pocket2Mol
b
TargetDiff
c
Lingo3DMol
d
PocketFlow
e
PMDM
f
MolCraft
g
UniLingo3DMol (W/O NCI & Anc.)
h
UniLingo3DMol (W/ NCI & Anc.)
MW
Generated molecules
Glide M
Glide D
QED
SAS

Supplementary Figure S3. Distributions of molecules generated by various models on the DUD-E *de novo* design evaluation set.

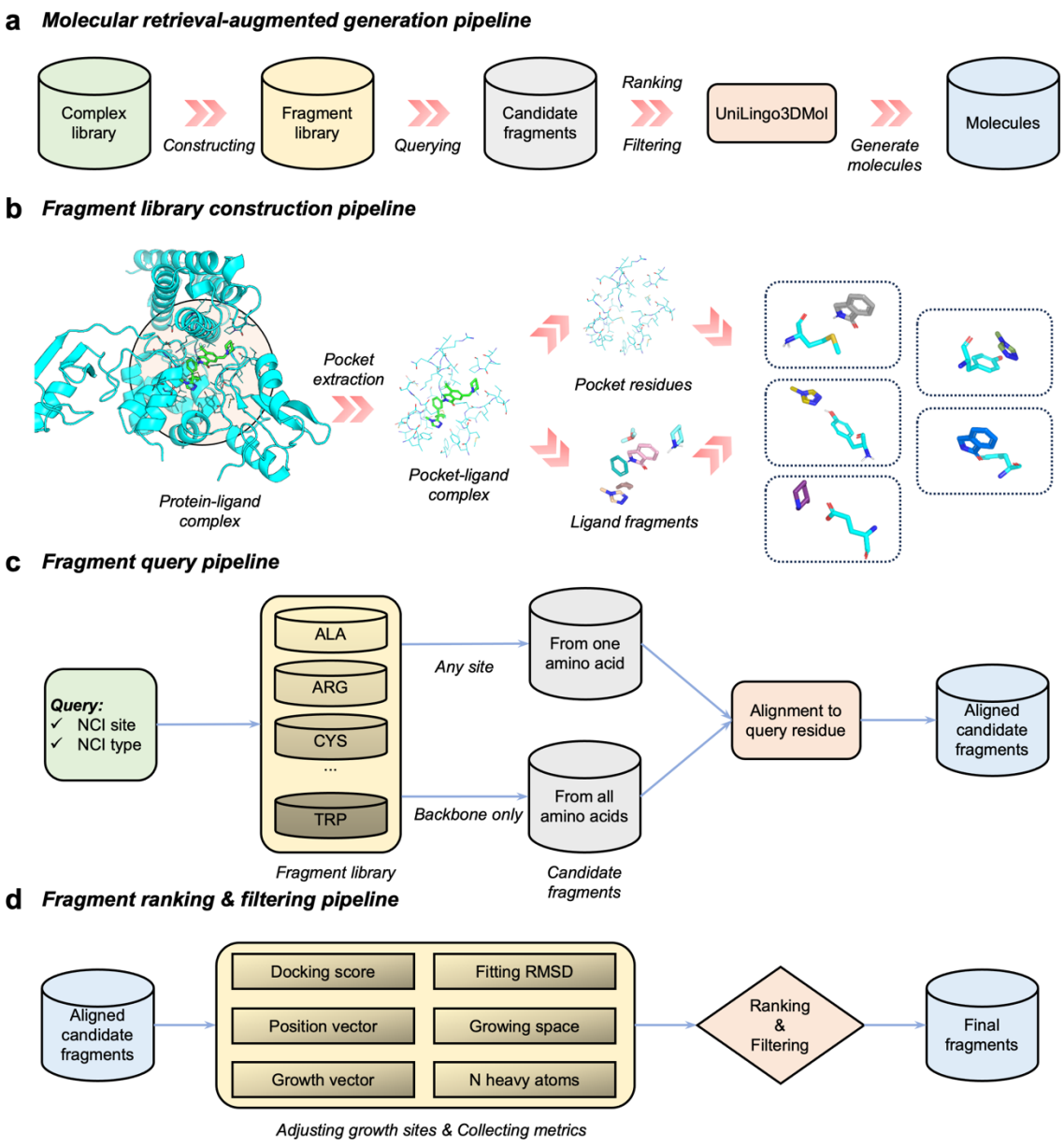
The region with QED > 0.3 and SAS < 5 is indicated with red boxes. **(a-h)** show the distributions on molecules generated by Pocket2Mol **(a)**, TargetDiff **(b)**, Lingo3DMol **(c)**, PocketFlow **(d)**, PMDM **(e)**, MolCraft **(f)**, UniLingo3DMol without NCI and anchor constraints **(g)**, and UniLingo3DMol with NCI and anchor constraints **(h)**, respectively. Heatmaps visualize the distribution of key properties across the generated molecules, including: (from left to right) molecular weight, counts, Glide M scores, and Glide D scores. These distributions are depicted along SAS and QED.



673

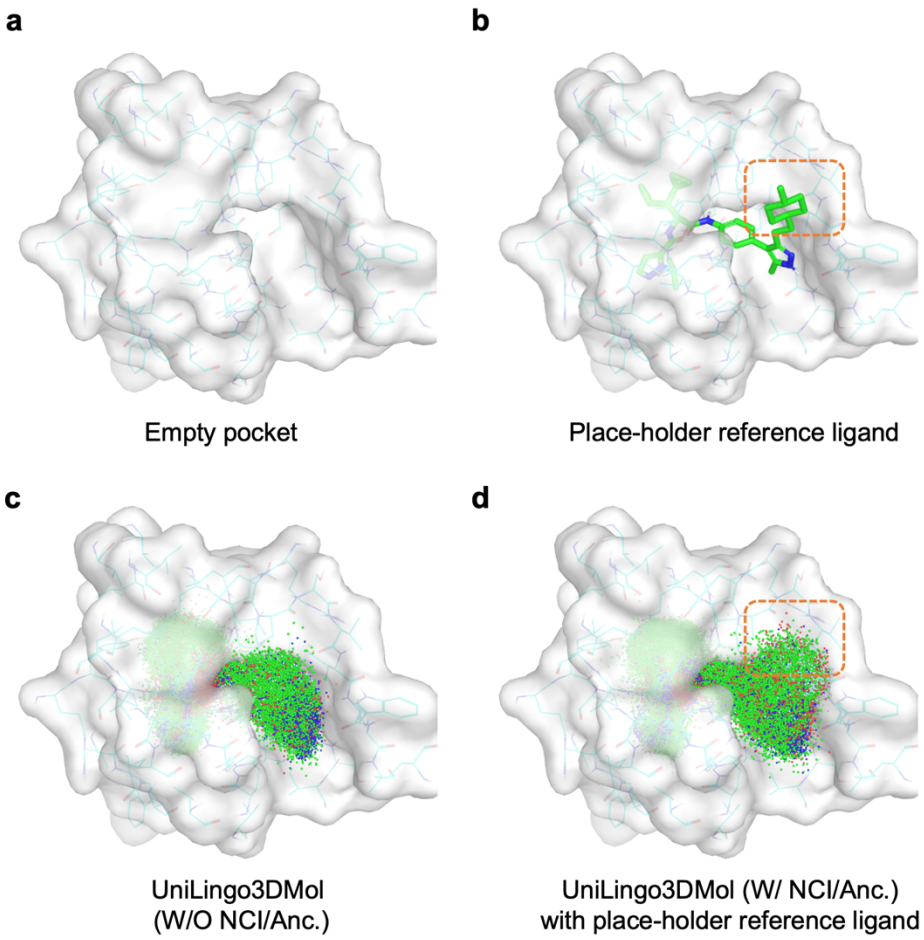
674 **Supplementary Figure S4. Distributions of molecules generated by various models on**
675 **the constrained design evaluation set.**

The drug-like region with QED > 0.3 and SAS < 5 is indicated with red boxes. **(a-i)** show distributions on the single **(a-c)**, two **(d-f)**, and three-fragment-retained **(g-i)** evaluation sets, respectively.



Supplementary Figure S5. Overview of MRAG schemes.

(a) Molecular retrieval-augmented generation pipeline.
(b) Fragment library construction pipeline.
(c) Fragment query pipeline.
(d) Fragment ranking and filtering pipeline.



Supplementary Figure S6. Shape control enabled by NCI/anchor conditioning in the binding pocket.

(a) Empty binding pocket without any reference structure for defining NCI/anchor constraints.
(b) Binding pocket with a reference ligand used as a spatial placeholder to define NCI and anchor constraints. The reference ligand does not need to be chemically meaningful and serves only to indicate the target region to be occupied; the moiety occupying this target region is highlighted by the orange dashed box.
(c) Spatial distribution of 500 generated molecules in the empty-pocket setting, visualized as an overlay of sampled molecules, with atoms represented as points.
(d) Spatial distribution of 500 generated molecules under NCI and anchor conditioning derived from the reference ligand, visualized as an overlay of sampled molecules. The generated molecules tend to extend toward and occupy the region highlighted by the orange dashed box, demonstrating that NCI and anchor conditioning can steer molecular generation toward the specified sub-pocket region. All results are based on PDB 7AMA, used here as a representative example. Molecules are generated via stochastic sampling,

and their superposition is used to illustrate the spatial distribution of generated molecules within the binding pocket under different conditioning settings.

4. Supplementary Tables S1-S10

Supplementary Table S1. Data collection, refinement and validation statistics.

	Cmpd. 7-CBL-B (PDB: 9WDH)
Data collection	
Space group	P21
Unit Cell	a=57.1 Å, b=102.0 Å, c=74.1 Å α=γ= 90°, β=89.86°
Resolution	2.8 Å (2.95 - 2.80)
Total measured reflections	20,806
Completeness (%)	99.1 (98.0)
Redundancy	3.2 (3.1)
I/σ	7.2 (2.1)
R_{merge}	0.172 (0.346)
CC(1/2)	0.988 (0.964)
Resolution range	59.9 – 2.8 Å
Refinement	
Initial model	PDB: 8GCY
R, R_{free} Model composition	25.4%, 30.7%
Non-hydrogen atoms	6,483
Protein atoms	6,224
Ligand atoms	144
Wilson B-factor ($Å^2$)	13.4
Average B factors ($Å^2$)	
Protein	15.9
Ligand	6.2
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.79
Validation	
MolProbity score	2.87
Poor rotamers (%)	2.3
Ramachandran plot	
Favored (%)	89.7
Allowed (%)	8.8
Disallowed (%)	1.5

Supplementary Table S2. Summary of ablation experiments.

Ablation experiment tables	Experimental contents
Supplementary Table S3	Ablation experiment of three-stage training strategy.
Supplementary Table S4	Ablation experiment of multi-task training strategy.
Supplementary Table S5	Ablation experiment of inference tasks.
Supplementary Table S6	Ablation experiment with/without fragment-retained data.
Supplementary Table S7	Ablation experiment of MRAG strategy.

Supplementary Table S3. Comparative ablation study of UniLingo3DMol-generated molecules under the three-stage training strategy. The top two results for each metric are highlighted in **bold** and underlined, separately.

	W/O Stage 1	W/O Stage 2	W/O Stage 3	Standard
# Generated molecules	97,116	102,000	102,000	101,439
MW	392±122	445±126	405±120	357±114
QED (↑)	0.52±0.21	0.40±0.21	0.49±0.21	**0.55±0.21**
SAS (↓)	**3.3±0.9**	4.0±1.0	3.5±1.0	3.4±1.0
RAS (↑)	**0.82±0.26**	0.66±0.34	0.79±0.29	0.80±0.29
# Drug-like molecules	68,360	47,470	66,122	72,421
% Drug-like molecules (↑)	70.4%	46.5%	64.8%	**71.4%**
Vina M (↓)	-4.1±1.9	-6.0±2.0	**-6.7±2.1**	**-6.7±1.9**
Vina D (↓)	-7.8±1.4	-8.0±1.3	**-8.2±1.5**	-7.9±1.5
Glide M (↓)	-5.0±1.9	-6.4±2.2	**-7.3±2.5**	-7.0±2.4
Glide D (↓)	-7.5±2.1	-7.3±1.8	**-8.0±2.2**	-7.8±2.1
The comparison below involves only drug-like molecules (QED > 0.3 and SAS < 5 and RAS >0.5)				
Topology properties				
MW	358±89	388±87	368±88	333±86
QED (↑)	0.59±0.16	0.55±0.16	0.58±0.16	**0.62±0.16**
SAS (↓)	**3.0±0.7**	3.3±0.7	3.1±0.7	**3.0±0.8**
RAS (↑)	**0.93±0.10**	0.90±0.12	0.92±0.11	**0.93±0.10**
Conformation qualities				
BL JSD (↓)	0.35±0.09	0.35±0.10	0.35±0.08	**0.34±0.08**
BA JSD (↓)	**0.23±0.07**	0.25±0.06	0.24±0.08	**0.23±0.07**
DA JSD (↓)	**0.14±0.10**	0.20±0.11	0.17±0.10	0.17±0.08
Binding modes				
Vina M (↓)	-4.1±2.0	-6.1±2.0	**-6.8±2.1**	-6.7±1.9
Vina D (↓)	-7.8±1.4	-8.0±1.4	**-8.2±1.5**	-7.8±1.5
Glide M (↓)	-5.0±1.9	-6.5±2.1	**-7.3±2.5**	-7.0±2.4
Glide D (↓)	-7.4±2.0	-7.3±1.8	**-8.0±2.2**	-7.7±2.0
% IMP (↑)	1.5%	8.7%	**13.3%**	10.2%
% NCI (↑)	3.1%	24.1%	33.8%	**36.0%**
Reproduction ability of active molecules				
% ECFP_TS > 0.5 (↑)	7.4%	15.7%	70.6%	**71.6%**

Note: Each of the above methods was evaluated on the *de novo* evaluation set by generating approximately 1,000 molecules without NCI and anchor site annotations.

Supplementary Table S4. Comparative ablation study of UniLingo3DMol-generated molecules under the multi-task training strategy. The top two results for each metric are highlighted in **bold** and underlined, separately.

	W/O Task 1	Standard
# Generated molecules	101,745	101,439
MW	364±115	357±114
QED (↑)	0.54±0.21	**0.55±0.21**
SAS (↓)	**3.5±1.0**	3.4±1.0
RAS (↑)	**0.80±0.29**	**0.80±0.29**
# Drug-like molecules	71,413	72,421
% Drug-like molecules (↑)	70.2%	**71.4%**
Vina M (↓)	-6.6±2.0	**-6.7±1.9**
Vina D (↓)	**-7.9±1.5**	**-7.9±1.5**
Glide M (↓)	-6.9±2.4	**-7.0±2.4**
Glide D (↓)	-7.7±2.1	**-7.8±2.1**
The comparison below involves only drug-like molecules (QED > 0.3 and SAS < 5 and RAS >0.5)		
Topology properties		
MW	339±88	333±86
QED (↑)	0.61±0.16	**0.62±0.16**
SAS (↓)	3.1±0.8	**3.0±0.8**
RAS (↑)	**0.93±0.10**	**0.93±0.10**
Conformation qualities		
BL JSD (↓)	**0.33±0.09**	0.34±0.08
BA JSD (↓)	**0.22±0.07**	0.23±0.07
DA JSD (↓)	**0.16±0.09**	0.17±0.08
Binding modes		
Vina M (↓)	-6.6±2.0	**-6.7±1.9**
Vina D (↓)	-7.9±1.5	-7.8±1.5
Glide M (↓)	-6.8±2.4	**-7.0±2.4**
Glide D (↓)	-7.6±2.1	**-7.7±2.0**
% IMP (↑)	9.0%	**10.2%**
% NCI (↑)	31.0%	**36.0%**
Reproduction ability of active molecules		
% ECFP_TS > 0.5 (↑)	70.6%	**71.6%**

Note: Each of the above methods was evaluated on the *de novo* evaluation set by generating approximately 1,000 molecules without NCI and anchor site annotations.

Supplementary Table S5. Comparative ablation study of UniLingo3DMol-generated molecules across inference tasks. The top result for each metric is highlighted in **bold**.

	W/ predicted NCI & Anc.	Standard
# Generated molecules	101,163	101,439
MW	418±119	357±114
QED (↑)	0.48±0.21	**0.55±0.21**
SAS (↓)	3.7±1.0	**3.4±1.0**
RAS (↑)	0.73±0.32	**0.80±0.29**
# Drug-like molecules	59,794	72,421
% Drug-like molecules (↑)	59.1%	**71.4%**
Vina M (↓)	**-6.8±2.0**	-6.7±1.9
Vina D (↓)	**-8.2±1.4**	-7.9±1.5
Glide M (↓)	**-7.4±2.5**	-7.0±2.4
Glide D (↓)	**-8.0±2.2**	-7.8±2.1
The comparison below involves only drug-like molecules (QED > 0.3 and SAS < 5 and RAS >0.5)		
Topology properties		
MW	380±84	333±86
QED (↑)	0.58±0.16	**0.62±0.16**
SAS (↓)	3.2±0.7	**3.0±0.8**
RAS (↑)	0.91±0.12	**0.93±0.10**
Conformation qualities		
BL JSD (↓)	0.36±0.09	**0.34±0.08**
BA JSD (↓)	**0.23±0.07**	**0.23±0.07**
DA JSD (↓)	0.18±0.09	**0.17±0.08**
Binding modes		
Vina M (↓)	**-6.9±2.1**	-6.7±1.9
Vina D (↓)	**-8.2±1.5**	-7.8±1.5
Glide M (↓)	**-7.4±2.5**	-7.0±2.4
Glide D (↓)	**-8.0±2.2**	-7.7±2.0
% IMP (↑)	**15.1%**	10.2%
% NCI (↑)	**42.4%**	36.0%
Reproduction ability of active molecules		
% ECFP_TS > 0.5 (↑)	57.8%	**71.6%**

Note: Each of the above methods was evaluated on the *de novo* evaluation set. The notation "predicted NCI and anchor" refers to a two-step process: first predicting NCI and anchor sites and then generating molecules with NCI and anchor site annotations.

Supplementary Table S6. Comparative ablation study of UniLingo3DMol-generated molecules with and without fragment-retained data. The top result for each metric is highlighted in **bold**.

	W/O fragment-retained data	Standard
# Generated molecules	102,000	101,439
MW	351±114	357±114
QED (↑)	**0.56±0.21**	0.55±0.21
SAS (↓)	**3.4±1.0**	**3.4±1.0**
RAS (↑)	**0.81±0.28**	0.80±0.29
# Drug-like molecules	73,748	72,421
% Drug-like molecules (↑)	**72.3%**	71.4%
Vina M (↓)	-6.5±2.0	**-6.7±1.9**
Vina D (↓)	-7.8±1.5	**-7.9±1.5**
Glide M (↓)	-6.8±2.4	**-7.0±2.4**
Glide D (↓)	-7.6±2.0	**-7.8±2.1**
The comparison below involves only drug-like molecules (QED > 0.3 and SAS < 5 and RAS >0.5)		
Topology properties		
MW	328±88	333±86
QED (↑)	**0.62±0.16**	**0.62±0.16**
SAS (↓)	**3.0±0.8**	**3.0±0.8**
RAS (↑)	**0.93±0.10**	**0.93±0.10**
Conformation qualities		
BL JSD (↓)	0.35±0.08	**0.34±0.08**
BA JSD (↓)	**0.22±0.07**	0.23±0.07
DA JSD (↓)	**0.17±0.09**	**0.17±0.08**
Binding modes		
Vina M (↓)	-6.4±2.0	**-6.7±1.9**
Vina D (↓)	**-7.8±1.5**	**-7.8±1.5**
Glide M (↓)	-6.7±2.3	**-7.0±2.4**
Glide D (↓)	-7.5±2.0	**-7.7±2.0**
% IMP (↑)	8.8%	**10.2%**
% NCI (↑)	31.0%	**36.0%**
Reproduction ability of active molecules		
% ECFP_TS > 0.5 (↑)	64.7%	**71.6%**

Note: Each of the above methods was evaluated on the *de novo* evaluation set by generating approximately 1,000 molecules without NCI and anchor site annotations.

Supplementary Table S7. Comparative ablation study of UniLingo3DMol-generated molecules under MRAG strategy. The top two results for each metric are highlighted in **bold** and underlined, separately.

	W/O MRAG		W/ MRAG	
	W/O NCI & Anc.	W/ NCI & Anc.	W/O NCI & Anc.	W/ NCI & Anc.
# Generated molecules	101,439	100,932	100,303	99,519
MW	357±114	362±113	395±112	393±112
QED (↑)	0.55±0.21	**0.56±0.20**	0.49±0.21	0.50±0.21
SAS (↓)	**3.4±1.0**	3.5±1.0	3.9±0.9	3.9±0.9
RAS (↑)	**0.80±0.29**	**0.80±0.29**	0.70±0.33	0.69±0.33
# Drug-like molecules	72,421	71,747	57,791	56,797
% Drug-like molecules (↑)	**71.4%**	71.1%	57.6%	57.1%
Vina M (↓)	-6.7±1.9	-6.7±2.0	**-6.9±2.0**	-6.8±2.0
Vina D (↓)	-7.9±1.5	-7.9±1.5	**-8.1±1.5**	**-8.1±1.5**
Glide M (↓)	-7.0±2.4	-7.2±2.5	**-7.6±2.5**	**-7.6±2.5**
Glide D (↓)	-7.8±2.1	**-7.9±2.1**	**-7.9±2.1**	**-7.9±2.2**
The comparison below involves only drug-like molecules (QED > 0.3 and SAS < 5 and RAS >0.5)				
Topology properties				
MW	333±86	337±86	356±84	356±84
QED (↑)	0.62±0.16	**0.63±0.16**	0.58±0.16	0.58±0.16
SAS (↓)	**3.0±0.8**	3.1±0.8	3.5±0.7	3.5±0.7
RAS (↑)	**0.93±0.10**	**0.93±0.10**	0.89±0.13	0.88±0.13
Conformation qualities				
BL JSD (↓)	0.34±0.08	0.33±0.09	**0.30±0.11**	0.34±0.07
BA JSD (↓)	0.23±0.07	0.23±0.07	**0.22±0.07**	0.23±0.07
DA JSD (↓)	**0.17±0.08**	0.18±0.08	0.19±0.06	0.19±0.05
Binding modes				
Vina M (↓)	-6.7±1.9	-6.7±2.0	**-6.9±2.0**	-6.8±2.0
Vina D (↓)	-7.8±1.5	-7.9±1.5	**-8.1±1.5**	-8.0±1.5
Glide M (↓)	-7.0±2.4	-7.2±2.4	-7.6±2.5	**-7.7±2.6**
Glide D (↓)	-7.7±2.0	-7.8±2.1	**-8.1±2.2**	**-8.1±2.2**
% IMP (↑)	10.2%	10.7%	12.3%	**13.6%**
% NCI (↑)	36.0%	43.1%	50.5%	**51.1%**
Reproduction ability of active molecules				
% ECFP_TS > 0.5 (↑)	71.6%	**74.5%**	29.4%	30.4%

Note: Each of the above methods was evaluated on the *de novo* evaluation set by generating approximately 1,000 molecules without NCI and anchor site annotations.

Supplementary Table S8. Valid vocabularies for atomic types, ring types, and DSMILES types.

Atomic type vocabulary					
<PAD>	<SOS>	<EOS>	<SEP>	C	N
O	S	P	F	Cl	Br
I	c	n	o	s	UnkElement
UnkAtomType	UnkSymbol	H	+	-	/
\	@	@@	#	=	1
2	3	4	5	6	(
)	[	]	[*]	([*])	<MOL>
Ring type vocabulary					
0	3	3+3	3+4	3+5	3+5+5
3+5+6	3+6	3+6+6	3+7	4	4+4
4+4+4	4+4+5	4+5	4+5+5	4+5+6	4+6
4+6+6	4+7	4+7+7	4+8	5	5+5
5+5+5	5+5+6	5+5+7	5+6	5+6+6	5+6+7
5+6+8	5+7	5+7+7	5+8	6	6+6
6+6+6	6+6+7	6+6+8	6+7	6+7+7	6+7+8
6+8	6+8+8	7	7+7	7+8	8
8+8	8+8+8				
DSMILES type vocabulary					
<PAD>	<SOS>	<EOS>	<SEP>	C_0	C_3
C_3+3	C_3+4	C_3+5	C_3+5+5	C_3+5+6	C_3+6
C_3+6+6	C_3+7	C_4	C_4+4	C_4+4+4	C_4+4+5
C_4+5	C_4+5+5	C_4+5+6	C_4+6	C_4+6+6	C_4+7
C_4+7+7	C_4+8	C_5	C_5+5	C_5+5+5	C_5+5+6
C_5+5+7	C_5+6	C_5+6+6	C_5+6+7	C_5+6+8	C_5+7
C_5+8	C_6	C_6+6	C_6+6+6	C_6+6+7	C_6+7
C_6+7+7	C_6+7+8	C_6+8	C_6+8+8	C_7	C_7+7
C_7+8	C_8	C_8+8	C_8+8+8	N_0	N_3
N_4	N_4+5	N_4+6	N_4+7	N_4+8	N_5
N_5+5	N_5+5+5	N_5+5+6	N_5+6	N_5+6+6	N_5+7
N_5+8	N_6	N_6+6	N_6+6+6	N_6+6+7	N_6+7
N_6+7+7	N_6+8	N_7	N_7+7	N_8	O_0
O_3	O_4	O_4+6	O_5	O_5+5	O_5+6
O_5+7	O_6	O_6+6	O_6+7	O_7	O_7+7

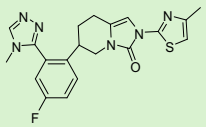
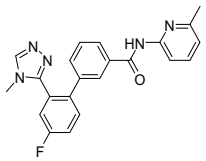
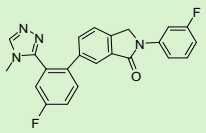
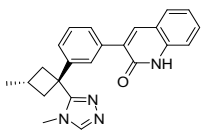
O_8	S_0	S_3	S_4	S_5	S_5+5
S_5+6	S_6	S_6+6	S_7	S_8	P_0
P_5	P_6	P_7	F_0	Cl_0	Br_0
I_0	c_0	c_3	c_4	c_4+6	c_5
c_5+5	c_5+5+6	c_5+6	c_5+6+6	c_5+6+7	c_5+6+8
c_5+7	c_5+7+7	c_5+8	c_6	c_6+6	c_6+6+6
c_6+6+7	c_6+6+8	c_6+8+8	c_6+7	c_6+7+7	c_6+8
c_7	c_7+7	n_0	n_5	n_5+5	n_5+5+6
n_5+6	n_5+6+6	n_5+6+7	n_5+7	n_5+7+7	n_5+8
n_6	n_6+6	n_6+6+6	n_6+7	n_6+7+7	n_6+8
n_7	o_5	o_6	s_5	s_6	UnkElement
UnkAtomType	UnkSymbol	H	<MOL>	+	-
/	\	@	@@	#	=
1	2	3	4	5	6
(	)	[	]	[*]	([*])

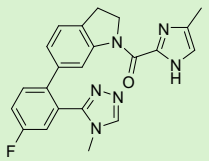
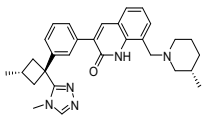
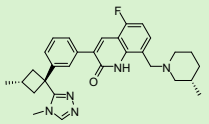
744

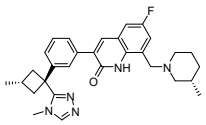
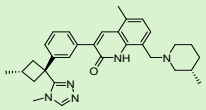
Supplementary Table S9. Hyperparameters in each stage of UniLingo3DMol.

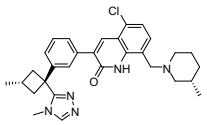
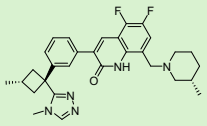
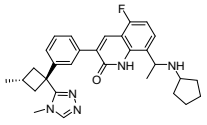
	Stage 1	Stage 2	Stage 3
Hyperparameter			
Number of encoder layers	6	6	6
Number of decoder layers	12	12	12
Max sequence length of encoder	150	500	500
Max sequence length of decoder	150	150	150
Hidden size	512	512	512
Number of attention heads	8	8	8
Intermediate size	1024	1024	1024
Number of head layers	2	2	2
Head hidden size	256	256	256
Training			
De novo data (include augmentation)	~405M	~58M	~1M
Fragment-retained data (include augmentation)	NA	~54M	~1M
Optimizer	AdamW	AdamW	AdamW
β1	0.9	0.9	0.9
β2	0.999	0.999	0.999
ε	1e-8	1e-8	1e-8
Learning rate	5e-5	5e-5	5e-5
weight decay	0.01	0.01	0.01
GPU type	H20	H20	H20
Number of GPUs	32	32	8
Hours to train	~226	~170	~0.5
Training epochs	500	200	10
Inference			
Sampling temperature (for *de novo* design)	NA	NA	0.5 (*suggested*)
Sampling temperature (for *fragment-retained* design)	NA	NA	1.5 (*suggested*)

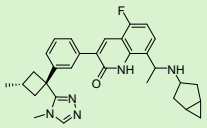
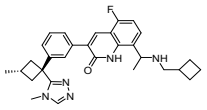
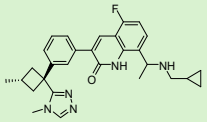
747 **Supplementary Table S10. Synthesized compounds.**

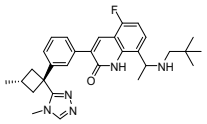
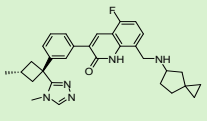
Cmpd. No	Structure	LCMS (ESI) m/z	^{1}H NMR
Cmpd. 1		411	^{1}H NMR (400 MHz, DMSO-d_6) δ 8.55 (s, 1H), 7.54 (dd, J = 8.7, 5.8 Hz, 1H), 7.44 - 7.39 (m, 1H), 7.34 -7.31 (m, 1H), 6.74 (d, J = 1.0 Hz, 1H), 6.37 (s, 1H), 4.33 (t, J = 9.4 Hz, 1H), 4.04 - 3.87 (m, 1H), 3.68 - 3.61 (m, 1H), 3.43 (s, 3H), 2.27 - 2.16 (m, 4H), 2.15 - 2.04 (m, 2H), 1.60 - 1.56 (m, 1H).
Cmpd. 2		388	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 9.39 (br s, 1H), 8.28 (d, J = 8.3 Hz, 1H), 8.02 (s, 1H), 8.00 (s, 1H), 7.91 (d, J = 7.8 Hz, 1H), 7.77 (t, J = 7.9 Hz, 1H), 7.67 - 7.61 (m, 1H), 7.42 - 7.34 (m, 3H), 7.26 - 7.22 (m, 1H), 7.01 (d, J = 7.5 Hz, 1H), 2.98 (s, 3H), 2.56 (s, 3H).
Cmpd. 3		403	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 8.00 (s, 1H), 7.93 (s, 1H), 7.79 (d, J = 11.3 Hz, 1H), 7.62 - 7.51 (m, 2H), 7.42 - 7.34 (m, 4H), 7.26 (s, 1H), 6.90 (t, J = 8.0 Hz, 1H), 4.83 (s, 2H), 3.00 (s, 3H).
Cmpd. 4		371	^{1}H NMR (400 MHz, DMSO-d_6) δ ppm: 11.92 (s, 1H), 8.30 (s, 1H), 8.09 (s, 1H), 7.78 - 7.71 (m, 2H), 7.66 - 7.58 (m, 1H), 7.55 - 7.46 (m,

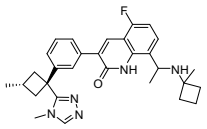
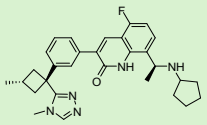
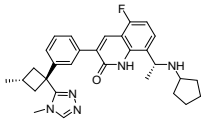
			1H), 7.42 (t, *J* = 7.7 Hz, 1H), 7.37 - 7.27 (m, 2H), 7.24 - 7.16 (m, 1H), 3.23 (s, 3H), 2.93 - 2.86 (m, 2H), 2.60 - 2.50 (m, 3H), 1.09 (d, *J* = 5.1 Hz, 3H).
Cmpd. 5		403	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 8.23 (s, 1H), 8.08 (s, 1H), 7.66 - 7.54 (m, 1H), 7.38 - 7.27 (m, 2H), 7.14 (d, *J* = 7.5 Hz, 1H), 6.92 (s, 1H), 6.87 (d, *J* = 7.1 Hz, 1H), 4.79 (t, *J* = 7.9 Hz, 2H), 3.21 (t, *J* = 8.3 Hz, 2H), 3.07 (s, 3H), 2.28 (s, 3H).
Cmpd. 6		482	^{1}H NMR (400 MHz, DMSO-*d6*) δ ppm: 12.08 (s, 1H), 8.28 (s, 1H), 8.14 (s, 1H), 7.80 (s, 1H), 7.69 (d, *J* = 7.2 Hz, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.50 - 7.25 (m, 3H), 7.16 (t, *J* = 7.6 Hz, 1H), 3.84 (s, 2H), 3.22 (s, 3H), 2.88 (d, *J* = 3.8 Hz, 2H), 2.78 (d, *J* = 9.4 Hz, 2H), 2.56 - 2.54 (m, 3H), 2.00 - 1.98 (m, 1H), 1.79 - 1.59 (m, 4H), 1.52 - 1.48 (m, 1H), 1.09 (d, *J* = 5.3 Hz, 3H), 0.95 - 0.93 (m, 1H), 0.85 (d, *J* = 6.2 Hz, 3H).
Cmpd. 7		500	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 9.35 (s, 1H), 8.11 (s, 1H), 8.09 (s, 1H), 7.82 (s, 1H), 7.58 (d, *J* = 7.6 Hz, 1H),

			7.44 (t, *J* = 7.7 Hz, 1H), 7.41 - 7.35 (m, 2H), 6.89 (t, *J* = 8.8 Hz, 1H), 4.01 (s, 2H), 3.30 (s, 3H), 3.06 - 2.88 (m, 4H), 2.68 - 2.55 (m, 3H), 2.22 (t, *J* = 10.9 Hz, 1H), 1.99 - 1.83 (m, 2H), 1.83 - 1.66 (m, 3H), 1.14 (d, *J* = 5.7 Hz, 3H), 1.05 - 0.92 (m, 1H), 0.88 (d, *J* = 6.2 Hz, 3H).
Cmpd. 8		500	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 12.34 (s, 1H), 7.98 (s, 1H), 7.81 (s, 1H), 7.77 (s, 1H), 7.62 (d, *J* = 7.4 Hz, 1H), 7.44 - 7.32 (m, 2H), 7.21 (dd, *J* = 8.2, 2.1 Hz, 1H), 7.08 (s, 1H), 3.83 (s, 2H), 3.26 (s, 3H), 2.93 - 2.86 (m, 4H), 2.67 (d, *J* = 6.1 Hz, 3H), 2.10 - 1.97 (m, 1H), 1.85 - 1.72 (m, 5H), 1.14 (d, *J* = 5.2 Hz, 3H), 0.98 - 0.89 (m, 4H).
Cmpd. 9		496	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 12.17 (s, 1H), 8.29 (s, 1H), 8.08 (s, 1H), 7.83 (s, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.44 (t, *J* = 7.7 Hz, 1H), 7.31 (d, *J* = 8.1 Hz, 1H), 7.25 (d, *J* = 7.5 Hz, 1H), 6.99 (d, *J* = 7.5 Hz, 1H), 3.79 (s, 2H), 3.24 (s, 3H), 2.93 - 2.83 (m, 2H), 2.83 - 2.72 (m, 2H), 2.60 - 2.52 (m, 6H), 2.04 - 1.93 (m, 1H), 1.76 - 1.59 (m, 4H), 1.56 - 1.42 (m, 1H), 1.09 (d, *J* = 5.2 Hz, 3H), 1.01

			- 0.90 (m, 1H), 0.85 (d, *J* = 6.2 Hz, 3H).
Cmpd. 10		516	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 12.49 (s, 1H), 8.29 (s, 1H), 8.10 (s, 1H), 7.75 (s, 1H), 7.59 (d, *J* = 7.7 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 1H), 7.41 - 7.35 (m, 2H), 7.30 (d, *J* = 7.9 Hz, 1H), 3.84 (s, 2H), 3.23 (s, 3H), 2.91 - 2.83 (m, 2H), 2.82 - 2.74 (m, 2H), 2.59 - 2.53 (m, 3H), 2.06 - 1.96 (m, 1H), 1.78 - 1.60 (m, 4H), 1.55 - 1.44 (m, 1H), 1.09 (d, *J* = 5.1 Hz, 3H), 0.99 - 0.91 (m, 1H), 0.85 (d, *J* = 6.3 Hz, 3H).
Cmpd. 11		518	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 8.03 (s, 1H), 7.99 (s, 1H), 7.86 (s, 1H), 7.58 (d, *J* = 7.5 Hz, 1H), 7.45 - 7.34 (m, 2H), 7.20 (s, 1H), 3.86 (s, 2H), 3.27 (s, 3H), 2.98 - 2.79 (m, 4H), 2.73 - 2.65 (m, 3H), 2.20 - 2.10 (m, 1H), 1.96 - 1.80 (m, 2H), 1.78 - 1.65 (m, 3H), 1.14 (d, *J* = 5.1 Hz, 3H), 1.03 - 0.95 (m, 1H), 0.87 (d, *J* = 5.5 Hz, 3H).
Cmpd. 12		500	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.25 (s, 1H), 7.98 (s, 1H), 7.72 (s, 1H), 7.57 (d, *J* = 7.7 Hz, 1H), 7.44 - 7.26 (m, 3H), 6.94 (t, *J* = 9.0 Hz, 1H), 4.23 - 4.14 (m, 1H), 3.19 (s, 3H), 2.88 - 2.75 (m,

			3H), 2.51 (d, *J* = 6.4 Hz, 3H), 1.80 - 1.50 (m, 4H), 1.44 - 1.22 (m, 7H), 1.05 (d, *J* = 4.8 Hz, 3H)
Cmpd. 13		512	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.28 (s, 1H), 8.02 (s, 1H), 7.75 (s, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.44 (t, *J* = 7.7 Hz, 1H), 7.40 - 7.27 (m, 2H), 7.07 - 6.91 (m, 1H), 4.26 - 4.08 (m, 1H), 3.23 (s, 3H), 3.18 - 3.09 (m, 1H), 2.93 - 2.82 (m, 2H), 2.60 - 2.50 (m, 3H), 2.18 - 2.06 (m, 1H), 2.06 - 1.92 (m, 1H), 1.48 (dd, *J* = 13.4, 4.4 Hz, 1H), 1.35 - 1.21 (m, 4H), 1.22 - 1.01 (m, 5H), 0.65 - 0.56 (m, 1H), 0.38 - 0.33 (m, 1H).
Cmpd. 14		500	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.24 (s, 1H), 7.99 (s, 1H), 7.70 (d, *J* = 1.9 Hz, 1H), 7.57 (d, *J* = 7.7 Hz, 1H), 7.43 - 7.26 (m, 3H), 6.94 (dd, *J* = 9.8, 8.2 Hz, 1H), 4.12 - 4.02 (m, 1H), 3.19 (s, 3H), 2.89 - 2.79 (m, 2H), 2.65 - 2.47 (m, 4H), 2.45 - 2.33 (m, 1H), 2.36 - 2.23 (m, 1H), 2.04 - 1.86 (m, 2H), 1.86 - 1.50 (m, 4H), 1.34 (d, *J* = 6.6 Hz, 3H), 1.05 (d, *J* = 5.1 Hz, 3H).
Cmpd. 15		486	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.24 (s, 1H), 7.98 (s, 1H), 7.70 (s, 1H), 7.57 (d,

			J = 7.7 Hz, 1H), 7.43 - 7.26 (m, 3H), 6.94 (t, *J* = 9.0 Hz, 1H), 4.20 - 4.10 (m, 1H), 3.19 (s, 3H), 2.89 - 2.79 (m, 2H), 2.56 - 2.47 (m, 3H), 2.42 - 2.33 (m, 1H), 2.19 - 2.10 (m, 1H), 1.35 (d, *J* = 6.6 Hz, 3H), 1.05 (d, *J* = 4.9 Hz, 3H), 0.90 - 0.79 (m, 1H), 0.43 - 0.27 (m, 2H), 0.09 - 0.03 (m, 2H).
Cmpd. 16		502	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 12.97 (s, 1H), 8.28 (s, 1H), 8.02 (s, 1H), 7.73 (s, 1H), 7.61 (d, *J* = 7.6 Hz, 1H), 7.43 (t, *J* = 7.7 Hz, 1H), 7.40 - 7.28 (m, 2H), 7.07 - 6.91 (m, 1H), 4.18 - 4.03 (m, 1H), 3.22 (s, 3H), 2.95 - 2.80 (m, 2H), 2.62 - 2.52 (m, 3H), 2.39 - 2.30 (m, 1H), 2.14 - 2.00 (m, 1H), 1.43 (d, *J* = 6.6 Hz, 3H), 1.08 (d, *J* = 5.1 Hz, 3H), 0.89 (s, 9H).
Cmpd. 17		512	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.28 (s, 1H), 8.03 (s, 1H), 7.74 (s, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.48 - 7.38 (m, 2H), 7.33 (d, *J* = 8.0, 1.4 Hz, 1H), 6.98 (dd, *J* = 9.9, 8.2 Hz, 1H), 4.04 (s, 2H), 3.22 (s, 3H), 3.21 - 3.13 (m, 1H), 2.92 - 2.82 (m, 2H), 2.60 - 2.52 (m, 3H), 2.01 - 1.88 (m, 1H), 1.75 - 1.42 (m, 5H), 1.09 (d, *J* = 5.0 Hz, 3H),

			0.53 - 0.42 (m, 2H), 0.41 - 0.30 (m, 2H).
Cmpd. 18		500	^{1}H NMR (400 MHz, $CDCl_3$) δ ppm: 8.10 (s, 1H), 8.07 (s, 1H), 7.89 (s, 1H), 7.58 (d, *J* = 7.4 Hz, 1H), 7.48 - 7.30 (m, 2H), 7.24 (s, 1H), 6.85 (t, *J* = 8.6 Hz, 1H), 4.42 - 4.30 (m, 1H), 3.29 (s, 3H), 2.96 (d, *J* = 5.9 Hz, 2H), 2.74 - 2.62 (m, 3H), 2.17 - 2.10 (m, 1H), 2.02 - 1.90 (m, 2H), 1.75 - 1.60 (m, 3H), 1.59 - 1.53 (m, 3H), 1.33 - 1.20 (m, 3H), 1.14 (d, *J* = 5.4 Hz, 3H)
Cmpd. 19		500	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.28 (s, 1H), 8.02 (s, 1H), 7.76 (s, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.48 - 7.29 (m, 3H), 6.98 (dd, *J* = 9.8, 8.3 Hz, 1H), 4.27 - 4.18 (m, 1H), 3.22 (s, 3H), 2.92 - 2.81 (m, 3H), 2.54 (d, *J* = 6.6 Hz, 3H), 1.83 - 1.73 (m, 1H), 1.72 - 1.53 (m, 3H), 1.50 - 1.26 (m, 7H), 1.09 (d, *J* = 4.9 Hz, 3H).
Cmpd. 20		500	^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ ppm: 8.28 (s, 1H), 8.02 (s, 1H), 7.76 (s, 1H), 7.61 (d, *J* = 7.8 Hz, 1H), 7.48 - 7.29 (m, 3H), 6.98 (dd, *J* = 9.8, 8.3 Hz, 1H), 4.27 - 4.18 (m, 1H), 3.22 (s, 3H), 2.92 - 2.82 (m, 3H), 2.54 (d, *J* = 6.5 Hz, 3H), 1.83 - 1.73 (m, 1H),

	1.72 - 1.53 (m, 3H), 1.50 - 1.25 (m, 7H), 1.09 (d, J = 4.8 Hz, 3H).

748

5. Supplementary Algorithms S1-S6

Algorithm S1 pre-training process

procedure LIGANDGENERATION($D_l^\dagger$, D_l)
    $\mathcal{L}_{\text{PT}} \leftarrow 0$
    Autoregressively predict with three ligand generation heads via teacher-forcing:

- Next token in DSMILES
- Atom pointer position
- Local & global coordinates

    Compute loss components:

$$\mathcal{L}_{\text{PT}} = \mathcal{L}_{\text{token}} + \mathcal{L}_{\text{ptr}} + \mathcal{L}_{\text{pos}}$$

    return $\mathcal{L}_{\text{PT}}$
end procedure

Require: Perturbed ligand data $D_l^\dagger$, ligand data represented by DSMILES D_l, total epochs E
Ensure: Model parameters
Shuffle data in dataloader
for epoch $e \leftarrow 1$ to E **do** $\mathcal{L}_{\text{PT}} \leftarrow$ LIGANDGENERATION($D_l^\dagger$, D_l)
    Backpropagate $\mathcal{L}_{\text{PT}}$
    Update parameters with AdamW optimizer
end for

Algorithm S2 Multi-task training process

procedure LIGANDGENERATION(D_p, D_l)
    $\mathcal{L}_\tau \leftarrow 0$
    Autoregressively predict with three ligand generation heads via teacher-forcing:

- Next token in DSMILES
- Atom pointer position
- Local & global coordinates

    Compute loss components:

$$\mathcal{L}_\tau = \mathcal{L}_{\text{token}} + \mathcal{L}_{\text{ptr}} + \mathcal{L}_{\text{pos}}$$

    return $\mathcal{L}_\tau$
end procedure

procedure NCIANCHORPREDICTION(D_p)
    $\mathcal{L}_\tau \leftarrow 0$
    Independently predict extra information with the pocket prediction head:

- NCI sites prediction
- Anchor sites prediction

    Compute loss components:

$$\mathcal{L}_\tau = \mathcal{L}_{\text{NCI}} + \mathcal{L}_{\text{Anchor}}$$

    return $\mathcal{L}_\tau$
end procedure

Require: Pocket data D_p, ligand data represented by DSMILES D_l, total epochs E
Ensure: Model parameters
Shuffle data in dataloader
Load pre-trained or post-trained parameters ▷ Depend on the training stage
for epoch $e \leftarrow 1$ to E **do**
    for task $\tau \in \{1, 2, 3\}$ **do** ▷ Multi-task iteration
        if $\tau = 1$ **then** ▷ Task1: NCI & Anc prediction
            $\mathcal{L}_\tau \leftarrow$ NCIANCHORPREDICTION(D_p)
        else if $\tau = 2$ **then** ▷ Task2: MG without NCI & Anc constraints
            $\mathcal{L}_\tau \leftarrow$ LIGANDGENERATION(D_p, D_l)
        else if $\tau = 3$ **then** ▷ Task3: MG with NCI & Anc constraints
            $\mathcal{L}_\tau \leftarrow$ LIGANDGENERATION(D_p, D_l)
        end if
        Backpropagate $\mathcal{L}_\tau$
        Update parameters with AdamW optimizer
    end for
end for

Algorithm S3 Random fragment sequence augmentation algorithm

Require: Original fragment sequence FS_{raw}
Ensure: Transformed fragment sequence FS_{trans}

$n \leftarrow |FS_{\text{raw}}|$ ▷ Obtain the number of fragments
$FS_{\text{trans}} \leftarrow [\,]$
while $n > 0$ **do**
    if FS_{trans} is empty **then**
        Randomly select fragment f from FS_{raw}
        $FS_{\text{trans}} \leftarrow$ Add f to FS_{trans}
        $FS_{\text{raw}} \leftarrow$ Remove f from FS_{raw}
    else
        Check **satisfied** connection constraints in FS_{trans}
        Randomly pick fragment f from FS_{raw} based on satisfied constraints
        $f_{\text{trans}} \leftarrow$ Use asterisk constraint to convert f
        $FS_{\text{trans}} \leftarrow$ Add f_{trans} to FS_{trans}
        $FS_{\text{raw}} \leftarrow$ Remove f from FS_{raw}
    end if
    $n \leftarrow n - 1$
end while
Verify the legality of the transformed fragment sequence FS_{trans}

Algorithm S4 Retained multiple fragments sequence augmentation algorithm

Require: Original fragment sequence FS_{raw}, number of retained fragments n_r
Ensure: Transformed fragment sequence FS_{trans}

$n \leftarrow |FS_{\text{raw}}|$ ▷ Obtain the number of fragments
$FS_{\text{trans}} \leftarrow [\,]$
if $n_{\text{r}} \geq n$ **then** ▷ Not meeting the need to preserve fragments
    return FS_{trans}
end if
$FS_{\text{trans}} \leftarrow [\,]$
while $n > 0$ **do**
    if FS_{trans} is empty **then**
        Randomly select fragment f from FS_{raw}
        $FS_{\text{trans}} \leftarrow$ Add f to FS_{trans}
        $FS_{\text{raw}} \leftarrow$ Remove f from FS_{raw}
        $n \leftarrow n - 1$
        $n_r \leftarrow n_r - 1$
        while $n > 0$ & $n_r > 0$ **do** ▷ Preserve fragments
            Check **unsatisfied** connection constraints in FS_{trans}
            Randomly pick fragment f from FS_{raw} based on unmet constraints
            $FS_{\text{trans}} \leftarrow$ Add f to FS_{trans}
            $FS_{\text{raw}} \leftarrow$ Remove f from FS_{raw}
            $n \leftarrow n - 1$
            $n_r \leftarrow n_r - 1$
        end while
    else
        Check **satisfied** connection constraints in FS_{trans}
        Randomly pick fragment f from FS_{raw} based on satisfied constraints
        $f_{\text{trans}} \leftarrow$ Use asterisk constraint to convert f
        $FS_{\text{trans}} \leftarrow$ Add f_{trans} to FS_{trans}
        $FS_{\text{raw}} \leftarrow$ Remove f from FS_{raw}
        $n \leftarrow n - 1$
    end if
end while
Verify the legality of the transformed fragment sequence FS_{trans}

Algorithm S5 Pointer prediction algorithm

Require: Current step i, current step's DSMILES token DT_i, current step's pointer logits z_i^{Ptr}, and historical pointers $\mathrm{Ptr}_{0:i-1}$

Ensure: Current step's pointer Ptr_i

if DT_i is connection site (',' [*] or ['*']) **then**
  $B, S \leftarrow \mathrm{shape}(z_i^{\mathrm{Ptr}})[0:2]$ ▷ Batch size and sequence length
  Initialize mask with minimum values: $m \leftarrow \mathrm{full}(-\infty, \mathrm{shape} = (B, S))$
  Find indices j of unpaired connection sites according to historical pointers $\mathrm{Ptr}_{0:i-1}$
  $m_j \leftarrow 0$ ▷ Unmask the above indices
  $z_i^{\mathrm{Ptr}} \leftarrow z_i^{\mathrm{Ptr}} + m$ ▷ Apply mask to logits
  $\mathrm{Ptr}_i \leftarrow \arg\max z_i^{\mathrm{Ptr}})$ ▷ Select best pointers
else
  $\mathrm{Ptr}_i = i$ ▷ Non connected site points to itself
end if
return Ptr_i

Algorithm S6 Molecular generation process

procedure GENERATEONESTEP(Action (i))
    Predict the $(i+1)^{\text{th}}$ DSMILES token
    Predict the $(i+1)^{\text{th}}$ pointer using the $(i+1)^{\text{th}}$ DSMILES token
    Predict local coordinates (r, θ, ϕ) and global coordinates (x, y, z) using the $(i+1)^{\text{th}}$ pointer and DSMILES token
    Define search space around predicted local coordinates for final token coordinates:

- Bond length: $r \pm 0.1$Å
- Bond Angle: $\theta \pm 2°$
- Dihedral Angle: $\phi \pm 2°$

    Find global coordinate with highest joint probability within search space
    Set final predicted coordinates for the $(i+1)^{\text{th}}$ DSMILES token
    return $(i+1)^{\text{th}}$ DSMILES token, pointer, and coordinates
end procedure

Set the maximum sequence length of the ligand to T
if single/multiple retained fragments exist **then**
    Initialize Action (s) based on single/multiple retained fragments
    Set the starting step $i_{\text{start}} \leftarrow s$
else
    Initialize Action (0) with $\langle \text{SOS} \rangle$
    Set the starting step $i_{\text{start}} \leftarrow 0$
end if
for each generation step i in $[i_{\text{start}}, T-1]$ **do**
    $\text{DT}_{i+1}, \text{Ptr}_{i+1}, \text{Pos}_{i+1} \leftarrow$ GENERATEONESTEP(Action (i)) ▷ Core step
    Update Action $(i+1)$ ▷ Update ligand embeddings
    if DT_{i+1} is $\langle \text{EOS} \rangle$ **then**
        break
    end if
end for

Reference

1. Zhou, G. *et al.* Uni-mol: A universal 3D molecular representation learning framework. in *International Conference on Learning Representations* (2023).
2. Jiao, R. *et al.* An equivariant pretrained transformer for unified 3D molecular representation learning. *Nat. Commun.* **17**, 2606 (2026).
3. Fan, F. *et al.* Assessing conformation validity and rationality of deep learning-generated 3D molecules. *Nat. Commun.* **17**, 2481 (2026).
4. Anstine, D. M., Zubatyuk, R. & Isayev, O. AIMNet2: A neural network potential to meet your neutral, charged, organic, and elemental-organic needs. *Chem. Sci.* **16**, 10228–10244 (2025).
5. Bouysset, C. & Fiorucci, S. ProLIF: A library to encode molecular interactions as fingerprints. *J. Cheminformatics* **13**, 72 (2021).
6. Greg Landrum *et al.* rdkit/rdkit: 2025_09_2 (Q3 2025) release. Zenodo https://doi.org/10.5281/ZENODO.591637 (2025).
7. He, X., Man, V. H., Yang, W., Lee, T.-S. & Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **153**, 114502 (2020).
8. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
9. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The cambridge structural database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
10. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
11. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology* vol. 276 307–326 (Elsevier, 1997).
12. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in phenix. *Biol. Crystallogr.* **75**, 861–877 (2019).