

An energy-based model empowers ultra-large virtual screening against 100-billion-scale chemical libraries

Shengyong Yang

yangsy@scu.edu.cn

Sichuan University <https://orcid.org/0000-0001-5147-3746>

Yuanyuan Jiang

Sichuan University

Chong Huang

Sichuan University

Mengzhe Dai

Sichuan University <https://orcid.org/0009-0006-5734-2812>

Rui Yao

Sichuan University

Weining Sun

Sichuan University

Liting Zhang

Sichuan University <https://orcid.org/0000-0003-2782-4424>

Qiao Huang

Sichuan University

Article

Keywords:

Posted Date: September 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10758423/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

An energy-based model empowers ultra-large virtual screening against 100-billion-scale chemical libraries

Yuanyuan Jiang^{1,2,3,#}, Chong Huang^{1,2,#}, Mengzhe Dai^{1,2,#}, Rui Yao^{1,2,#}, Weining Sun^{1,2,#}, Liting Zhang^{1,2,#}, Qiao Huang^{1,2}, Shengyong Yang^{1,2,*}

¹ Department of Biotherapy, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, Sichuan, 610041, China.

² New Cornerstone Science Laboratory, West China Hospital, Sichuan University, Chengdu, Sichuan, 610041, China.

³ Center for Natural Products Research, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan, 610041, China.

*To whom correspondence should be addressed.

These authors contributed equally to this work.

Corresponding author: Shengyong Yang, Email: yangsy@scu.edu.cn.

Abstract

Recent advances have expanded the accessible chemical space to nearly 100 billion molecules [1, 2], offering unprecedented opportunities for discovering novel chemotypes in drug development. However, screening such vast libraries using conventional methods remains prohibitively expensive, thus inaccessible to most drug developers [3]. To address this challenge, we propose DrugJEPA, an energy-based model that integrates contrastive learning with a Joint Embedding Predictive Architecture (JEPA). DrugJEPA achieves state-of-the-art accuracy on computational benchmarks. Furthermore, we construct a tailored virtual library of approximately 100 billion compounds (ZEUS-3D). Using this library, DrugJEPA completes a virtual screening campaign in just 28 hours on four GPUs—representing a speedup of up to seven orders of magnitude compared to molecular docking. In wet-laboratory validation experiments targeting three distinct protein classes (TYK2, TAAR1, and PUS1, representing kinase, GPCR, and epigenetic targets, respectively), DrugJEPA achieved high hit rates of 40%, 20%, and 30%. Notably, for some of the established targets, DrugJEPA identified single-digit nanomolar active molecules, while for entirely new targets, it still offered nanomolar-potency hits. This work represents a pioneering effort in ultra-fast virtual screening of ultra-large chemical libraries, making such screening affordable for conventional drug development programs.

Introduction

The discovery of active compounds from known chemical libraries—whether through physical or virtual screening—remains a mainstream strategy for identifying initial hits in drug development [4-7]. However, conventional screening approaches can only explore limited chemical space due to their technological and cost constraints, thereby limiting the discovery of novel, high-quality hits. In physical high-throughput screening (HTS), for example, the high costs of compound handling and screening equipment restrict practical testing to around 100,000 compounds per typical campaign [8]. Virtual screening (VS) employs computational models to rapidly identify compounds likely to bind a target of interest, allowing exploration of larger chemical libraries. However, traditional VS methods, such as molecular docking, remain computationally expensive and time-consuming due to the need to simulate protein–ligand interactions, restricting typical affordable projects to screening libraries on the order of 100 million compounds [9-13].

Expanding library size is generally considered beneficial for increasing the probability of discovering high-quality active compounds with new scaffolds [1, 3]. Consequently, considerable efforts have been made to enlarge the chemical space accessible for screening [14], with recent advances reaching nearly 100 billion molecules (so-called ultra-large libraries) [2]. However, this scale lies beyond the capabilities of conventional screening methods and remains inaccessible. Emerging artificial intelligence (AI) offers great potential for affordable, fast high-throughput screening, yet existing approaches struggle to balance speed and accuracy when handling ultra-large libraries [15]. Some methods, such as cascade screening, employ sequential filtering stages to narrow candidate pools but risk overlooking active compounds and fail to evaluate all molecules uniformly [16, 17]. Furthermore, storing and processing such massive libraries poses serious infrastructural challenges. For example, preparing 100 billion molecules for VS would require approximately 300 TB of storage to hold their 3D structures, incurring substantial costs and input/output (I/O) limitations.

Contrastive learning [18, 19], an emerging machine learning technique, has recently been applied to virtual screening and achieves a favorable trade-off between speed and accuracy [20-22]. Nevertheless, most contrastive learning models focus on global semantic alignment, often neglecting fine-grained features. This limitation can prevent the model from recognizing subtle differences in ligand functional groups or local changes in protein active sites that affect binding, leading to incorrect predictions. Considering the limitations of purely contrastive approaches, we shift to the Energy-Based Model (EBM) paradigm [23-26]. In an EBM, a learnable scalar energy function $E(x)$ or $E(x,y)$ quantifies the compatibility of a configuration, mapping positive samples onto a low-energy manifold. From an energy perspective, ligand–protein binding is fundamentally a thermodynamics-driven physical process, whose stability is governed by the free energy change $\Delta G = \Delta H - T\Delta S$. The internal energy term ΔH captures the strict geometric and interaction complementarity between the binding partners, which formally aligns with the compatibility objective of the InfoNCE loss. However, InfoNCE loss neglects the critical entropy term ($-T\Delta S$). In realistic drug

screening scenarios, this entropy term embodies the uncertainty inherent in the binding process, such as the conformational flexibility of the ligand in solution and the side-chain rearrangements within the binding pocket. This unmodeled uncertainty can limit the resolution of models relying solely on contrastive learning when distinguishing true tight-binding conformations from spurious weak-binding poses, and such models may consequently fall short of accuracy in ULVS.

To overcome this, we propose DrugJEPA, a new EBM that integrates contrastive learning with a Joint Embedding Predictive Architecture (JEPA) [27-30] and Mixture-of-Experts (MoE) design [31-34]. By introducing a JEPA-style $L2$ prediction loss, the model learns to predict the ligand representation from the pocket representation in an abstract embedding space, and vice versa. This bidirectional predictability constraint compels the model to discard irrelevant cross-modal noise and extract causal features, thereby functionally modeling and reducing the inter-modal uncertainty during binding—analogous to the entropy term in thermodynamics. Ultimately, the contrastive loss underpins the internal energy basis of binding, while the JEPA prediction loss suppresses uncertainty; working in concert, they enable DrugJEPA to more sharply separate genuine binding modes from decoy false positives on the energy landscape, substantially improving the discrimination of ligand–protein complexes. Furthermore, we also consider the biological reality—where proteins often bind multiple ligands—we further adopt multi-label contrastive learning, which captures shared binding modes across chemically diverse compound sets. Architecturally, we introduce an upcycling strategy [35] that injects sparsity into the encoder, transforming it into a Mixture-of-Experts (MoE) system. This expands model capacity and representational diversity with minimal additional inference cost, yielding stronger generalization and robustness. Furthermore, by incorporating protein sequence information, we embed rich evolutionary context into the model, significantly enhancing its early enrichment capability.

DrugJEPA achieves state-of-the-art performance on two widely used virtual screening benchmarks, DUD-E and LIT-PCBA, outperforming traditional docking-based methods and other deep neural networks. We then constructed a virtual chemical library of approximately 100 billion compounds. To improve I/O performance, reliability, and data capacity, we adopted RAID 5 technology with a binary format similar to in-memory representation. With this library, DrugJEPA completes a virtual screening campaign on just four GPUs in only 28 hours. To further validate its real-world utility, we applied DrugJEPA to screen small molecule ligands against three distinct protein targets—Tyk2 (kinase), TAAR1 (G-protein-coupled receptor), and PUS1 (epigenetic regulation protein)—using the 100-billion-scale library. From these screens, we obtained nanomolar-active compounds for all targets, with hit rates of 40%, 20%, and 30%, respectively. Collectively, these results demonstrate that DrugJEPA is an effective, ultra-fast, and ultra-large-scale virtual screening tool affordable for most drug developers.

Results

Design of the DrugJEPA model

DrugJEPA is a multimodal neural network designed for ultra-large-scale virtual screening. It employs contrastive learning to enhance feature representation and improve class separability between active and inactive molecules. Notably, this model integrates the Joint Embedding Predictive Architecture (JEPA) [27-30]. JEPA shifts focus from reconstructing raw inputs to predicting relationships within a shared embedding space, which facilitates more efficient representation learning by capturing high-level semantic patterns rather than low-level details. Moreover, to enhance efficiency, scalability, and performance of this model, DrugJEPA also incorporates the Mixture of Experts (MoE) architecture [31-34]. MoE activates only the most relevant sub-models (experts) for a given input, drastically reducing computational overhead compared to traditional dense models.

In the proposed DrugJEPA framework, we first encode protein pockets and small-molecule ligands into a shared unified latent representation space, facilitating cross-modal feature alignment at the latent level (Fig. 1a). To perform effective contrastive learning, we deploy two task-specific mappers to project molecular and pocket embeddings into a space of identical dimension, respectively (black lines in Fig. 1a). Notably, DrugJEPA incorporates a dedicated bidirectional prediction paradigm for cross-modal modeling: a molecule predictor is tasked with predicting the latent embeddings of molecule from the latent representation of pocket, whereas a pocket predictor is trained to infer the latent embeddings of pocket based on the molecular latent representation (gray dashed lines in Fig. 1a). A contrastive learning objective aligned with multiple positive samples (Fig. 1b) serves as the primary training signal, bringing interacting pocket–ligand pairs closer and pushing non-interacting ones apart, while explicitly accounting for the biological fact that a single pocket can have multiple ligands. This enables the model to construct a latent space with strong semantic consistency and enhanced discriminative capability. To further enhance representation quality, we incorporate a JEPA-based regularization mechanism (Fig. 1c), which performs cross-modal prediction in the latent space (e.g., ligand→pocket and pocket→ligand). Instead of reconstructing raw structures, this objective enables the model to predict one modality from the other, which explicitly encourages pockets and ligands to form mutually predictable representations, thereby pulling geometrically compatible embeddings closer in the latent space.

To enhance the representation ability of fine-grained features, DrugJEPA additionally introduces a ranking loss [21, 36] (Fig. 1d), which enforces ligands with higher experimental affinity to be ranked more favorably in the latent space or among predicted scores. In this way, ranking accuracy is directly optimized during training rather than treated as a post-hoc scoring procedure. Furthermore, protein sequence information is integrated through contrastive and ranking supervision (Fig. 1a), enhancing the latent space with evolutionary constraints and specificity signals that are difficult to recover solely from pocket structures.

To further enhance representational capacity, we designed a novel encoder architecture based on the MoE framework (Fig. 1(e-g)). Specifically, we adopted an

upcycling strategy (Fig. 1f) [35] that converts the dense transformer encoder into a sparse architecture, replacing the conventional MLP block with a MoE block whose parameters are initialized by reusing those from the pretrained dense model. Rather than upcycling all dense layers, we introduce an alternating stacking scheme of sparse–dense–sparse layers [37], which balances efficiency and expressiveness. In this work, the upcycling strategy is applied to the pocket and ligand encoders, while the sequence encoder remains the pretrained ESMC model [38].

At the inference stage, we precompute and locally store the embeddings of molecules and pockets. Cosine similarity between molecule and pocket embeddings is then computed and ranked, thereby enriching the top-ranked candidates. Compared with conventional docking methods, DrugJEPA accelerates traditional virtual screening by up to seven orders of magnitude (see Supplementary Methods).

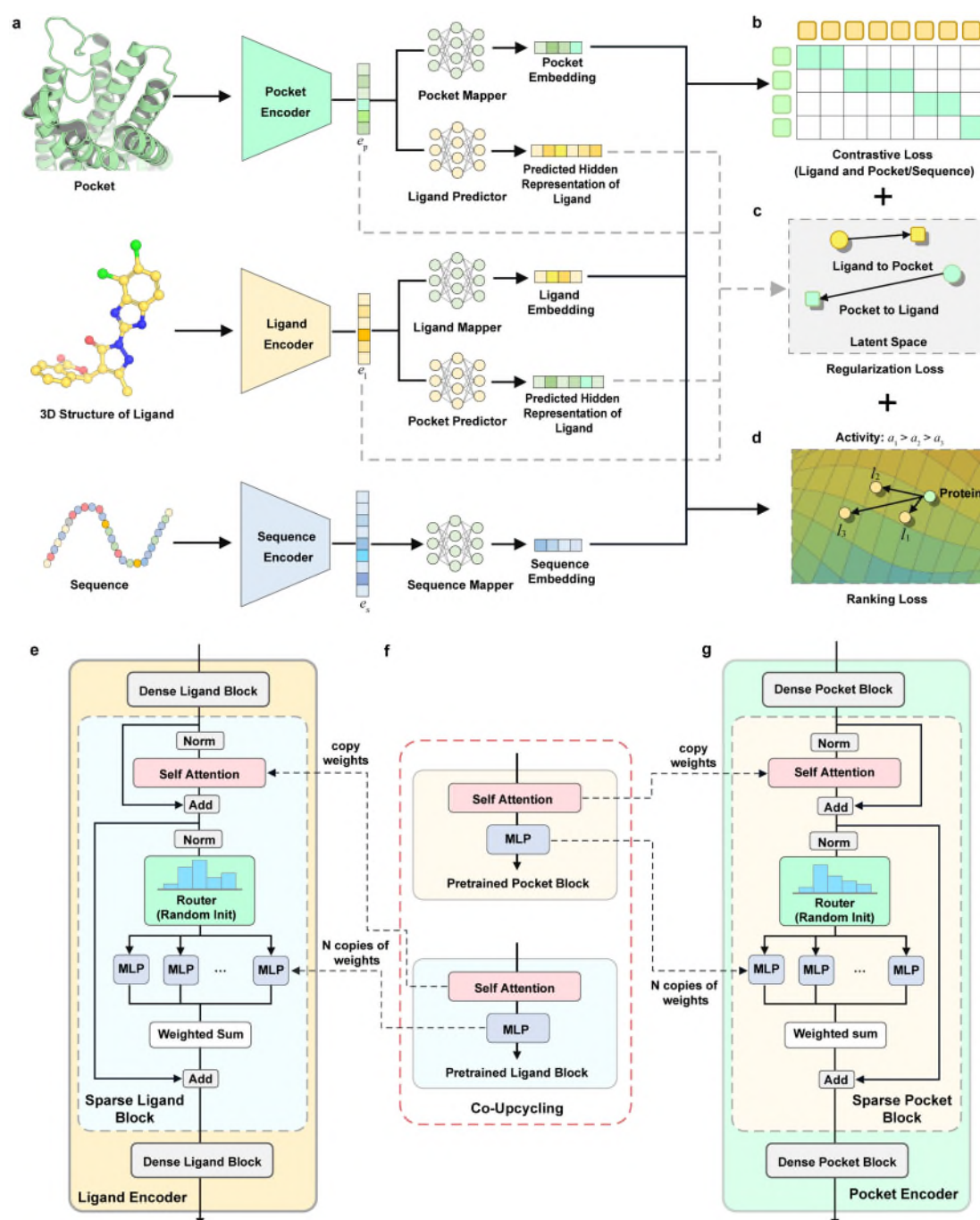


Figure. 1 The architecture of DrugJEPA. (a) Overview of the model architecture. DrugJEPA receives pocket, 3D molecule and sequence as input, and returns the embedding (e_p , e_l , e_s) for each input. These embeddings are converted into corresponding representations via the respective mapper and predictor, which are used for the calculation of contrastive loss, regularization loss, and ranking loss. (b) The multilabel contrastive loss encourages each pocket representation to align simultaneously with multiple ligand molecules, minimizing the angular distance between matching pocket and ligand embeddings. (c) The regularization loss constrains ligand embeddings that conform to the binding pattern to reside in close proximity to the corresponding pocket embeddings within the latent space. (d) The ranking loss guides the model to assign pocket-aware distances in the latent energy landscape,

enabling it to discriminate between distinct active ligands. a_i is the activity of ligand l_i . (e) The MoE architecture of ligand encoder. (f) A co-upcycling strategy is used for initializing trainable weights from pretrained models. (g) The MoE architecture of pocket encoder.

Evaluation of DrugJEPA on DUD-E and LIT-PCBA

For assessing the performance of DrugJEPA on virtual screening, we apply it to two classic datasets, DUD-E [39] and LIT-PCBA [40]. Three classic metrics were used to evaluate virtual screening performance: AUROC (Area Under the Receiver Operating characteristic Curve), BEDROC (Boltzmann-Enhanced Discrimination of ROC Discrimination of ROC) and EF (Enrichment Factor). AUROC quantifies a model's discriminatory power to rank active ligands above inactive ones. BEDROC uses exponential weighting to prioritize early active compound recognition. EF-1% measures active enrichment in the top 1% of ranked ligands.

DUD-E is the most extensively used benchmark in virtual screening, which adopts a property-matched active-decoy pairing strategy to eliminate evaluation bias, allowing for robust assessment of the model's active ligand discrimination performance. As shown in Fig. 2(a-c) and Supplementary Table 1, traditional docking methods (blue in Fig. 2(a-c)), including Glide-SP and Vina, achieved moderate AUC values (71~77) but exhibited limited early enrichment on DUD-E. Classic deep learning-based approaches (orange in Fig. 2(a-c)) also performed poorly, with AUC values ranging from 50 to 68 and enrichment factor scores close to random. By contrast, contrastive learning-based methods (green in Fig. 2(a-c)) yielded substantial improvements. DrugCLIP [20] and FragCLIP [41] significantly outperformed conventional baselines, while LigUnity_{pocket} [21] and DrugHash [22] achieved even higher performance. Notably, DrugJEPA delivered the best overall results, attaining the highest AUC (89.61), BEDROC (71.13), and EF-1% (47.03) across all thresholds, highlighting its superior ability to prioritize active ligands at the earliest ranks.

LIT-PCBA is a more challenging dataset, as all its data originate from real screening experiments. As shown in Fig. 2(d-f) and Supplementary Table 2, overall performance dropped markedly compared to that on DUD-E, reflecting LIT-PCBA's reduced bias and higher complexity. Traditional docking methods performed near randomly, while classical deep learning models offered only marginal improvements. Nevertheless, representation learning models maintained relative advantages, with DrugCLIP, LigUnity_{pocket}, DrugHash, and DrugJEPA all achieving EF-1% values around 5-6, substantially higher than traditional baselines. Although its AUC (57.93) was not outstanding, DrugJEPA attained the top EF-1% (6.45) and BEDROC (7.33), indicating strong enrichment of active ligands in the top 1% and reliable early-hit retrieval under realistic conditions.

To further investigate the spatial distribution of molecular and pocket embeddings, we employed Uniform Manifold Approximation and Projection (UMAP) [42, 43] for visualization, using DUDE as an example. As shown in Fig. 2g, embeddings of active molecules in the DUDE dataset exhibited substantial overlap with those of the pockets. Four targets from DUDE—MMP13, DEF, AA2AR and ADA—were selected for

detailed illustration. Magnification of representative regions of these four targets further revealed that active molecules predominantly clustered near the pocket embeddings. This observation substantiates the promising early enrichment capability of DrugJEPA in virtual screening.

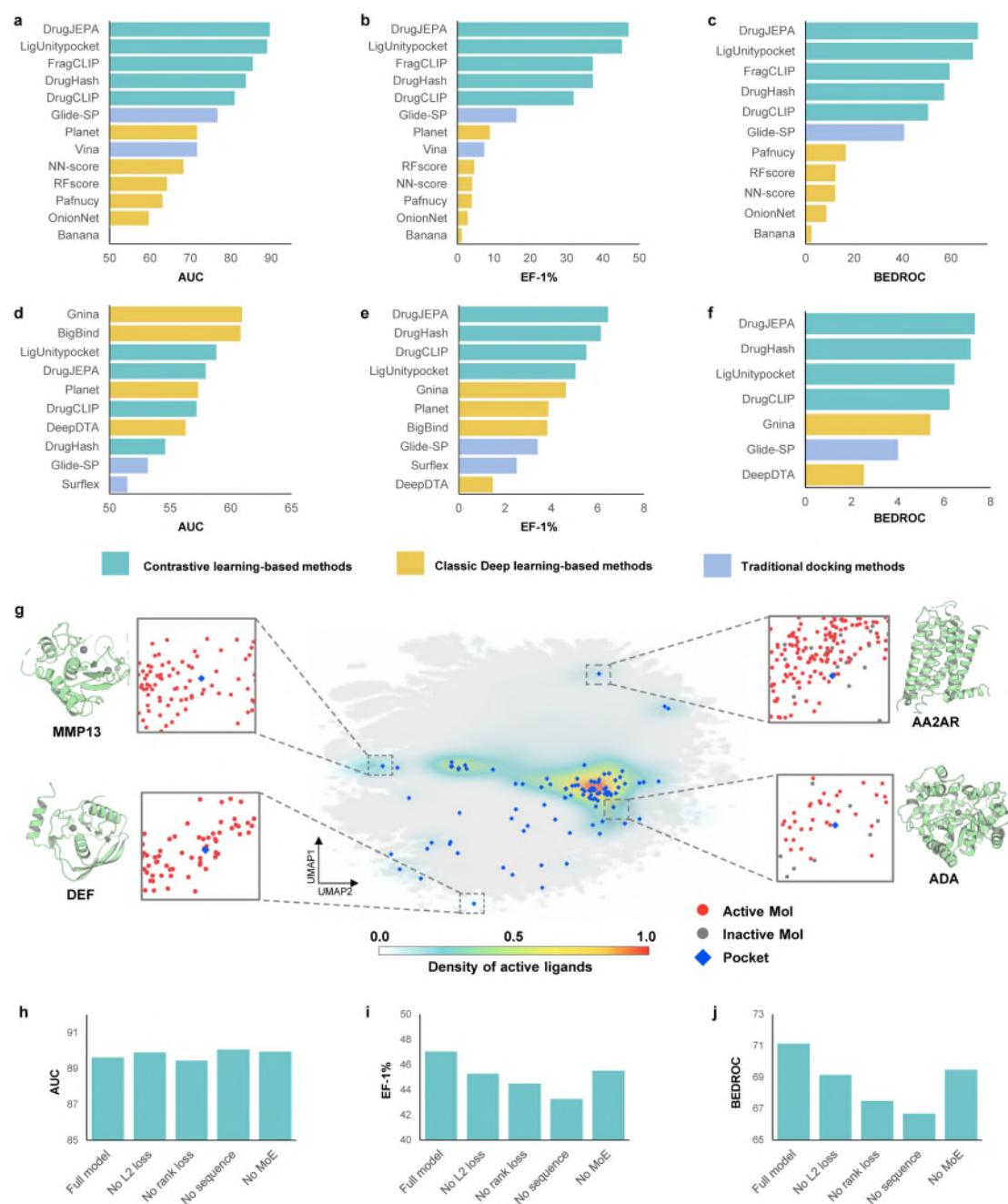


Figure 2. Results of dry-lab experiments. (a-c) Performance comparison on the DUD-E dataset for AUC, BEDROC, and EF-1%, respectively. (d-f) Corresponding results on the LIT-PCBA dataset. Green, orange, and blue bars represent traditional docking methods, classic deep learning-based methods, and contrastive learning-based methods, respectively. (g) UMAP visualization of ligand and pocket embeddings from the DUD-E dataset. The gray region indicates the projection of inactive molecule embeddings. Four representative DUD-E targets—MMP13, DEF, AA2AR, and ADA—

are highlighted as examples. (h-j) Results of ablation experiments.

Ablation Experiments

To verify the contribution of each component in the proposed model to its performance, we conducted ablation experiments by sequentially removing key modules, including JEPA-based L2 regularization loss, rank loss, sequence module, and mixture of experts (MoE). As shown in Fig. 2(h-j), the results of full model serve as the baseline for comparison. Notably, removing L2 regularization led to a decrease in BEDROC (69.145) and EF-1% (45.272), despite a marginal increase in AUC. This suggests that L2 regularization plays an important role in maintaining the model's enrichment capabilities. When rank loss was excluded, both BEDROC and EF-1% dropped significantly to 67.484 and 44.478, respectively. Although AUC remained relatively stable, these declines underscore the importance of rank loss for preserving the correct ordering of molecular candidates. Rank loss is useful when ensuring that the top-ranked candidates are correctly identified. Interestingly, removing sequence data led to the highest AUC (90.050), but a notable decrease in BEDROC (66.668) and EF-1% (43.258). This result indicates that evolutionary information of sequence data is vital for enhancing the enrichment capacity of model, even if it results in a slight sacrifice in classification accuracy. Finally, the exclusion of the Mixture of Experts (MoE) component reduced AUC (89.928), BEDROC (69.462) and EF-1% (45.502). This suggests that MoE, while not drastically affecting AUC, plays a role in enhancing ranking and enrichment performance, likely by allowing the model to specialize and better handle different subsets of the data.

Overall, the ablation results demonstrate that the rank loss, MoE, regularization loss, and sequence data act complementarily to improve early enrichment without sacrificing overall accuracy. These findings support a design principle that combines moderate architectural complexity, targeted regularization, and rank-oriented training to enhance practical large-scale virtual screening.

Embedding generation and screening platform establishment

To enable efficient VS against ultra-large-scale chemical libraries, we precomputed embeddings for the three-dimensional (3D) structures of approximately 100 billion small molecules. As illustrated in Fig. 3a, these molecules were sourced from on-demand virtual chemical libraries, including ZINC20 [44], ZINC22 [2], and Enamine Real [45]. The initial set comprised roughly 101 billion two-dimensional (2D) molecules, which were converted to 3D using RDKit [46] with the ETKDGV3 algorithm [47]. For each molecule, up to five representative conformations were generated, resulting in a database named ZEUS-3D (ZINC-Enamine Unified Structures of 3D). The final ZEUS-3D database contains approximately 96.4 billion 3D molecules with over 450 billion conformations, occupying 156 TB of storage (Supplementary Tables 3-4). Conformation generation was performed in parallel on a computer cluster, consuming about 10 million core-hours over four months. To our knowledge, ZEUS-3D is the largest 3D molecular database currently available.

Using ZEUS-3D, we computed embedding representations for each conformation

with DrugJEPA (Fig. 3a). Given the hundreds of billions of embeddings, data access efficiency became the primary bottleneck during screening. To address this, we implemented a binary storage format optimized for sequential reading in large-scale retrieval tasks. Embeddings were stored alongside molecular identifiers, SMILES strings, and associated metadata in sharded local files. This approach minimized storage overhead and I/O load, enabling efficient library access on conventional servers. The total storage for all virtual screening data reached 114 TB. To further improve I/O performance and safeguard against data loss, we employed a Redundant Array of Independent Disks (RAID) configuration using RAID5. Additional details are provided in Supplementary Methods.

For protein pocket embedding generation, we distinguished between novel and known proteins. For a novel protein, we used predicted structures from AlphaFoldDB, retaining only residues with high confidence scores ($pLDDT \geq 70$). Pockets were then detected using Schrödinger software (or equivalent tools). For proteins with structures in the Protein Data Bank (PDB) [48], we retrieved all available structures from both the PDB and AlphaFoldDB [49] and aligned them to the corresponding AlphaFold reference structure. Ligands from reported structures guided pocket extraction by selecting residues within 6 Å of any ligand heavy atom. For each protein, all extracted pockets were clustered to identify distinct binding sites. Finally, pocket embeddings were computed using the pocket encoder of DrugJEPA. To facilitate broad usage, we constructed a UniProt [50] indexed library of protein pockets, termed DJ-Pockets, based on both experimental (PDB) and predicted (AlphaFoldDB) structures (Fig. 3b). This library includes over 170,000 pockets with experimental structures and 17,000 with predicted structures. Subsequently, an embedding library, named DJ-Pockets-emb, was generated using the pocket encoder module of DrugJEPA. Details of the construction of DJ-Pockets and DJ-Pockets-emb are described in the Supplementary Methods section, Supplementary Fig. 1, and Supplementary Tables 5-7.

The overall screening process comprises two stages: ultra-high-throughput virtual screening and post-processing refinement. In the first stage, we calculated the cosine similarity between each molecule and the target pocket embeddings (Fig. 3c, left). All pocket conformations associated with a given protein were considered jointly, implicitly accounting for protein flexibility and enabling screening across a conformational ensemble that better reflects biological reality. For each ligand, the highest similarity score across all pocket-ligand conformer pairs was retained. Asynchronous parallel scheduling was adopted to overlap data loading and GPU computing processes, further enhancing computational efficiency. Using a server equipped with low-cost hard disk drives (HDDs) and four consumer-grade GPUs, screening the entire ZEUS-3D library against multiple protein targets was completed in approximately 28 hours—equivalent to scoring over one quadrillion pocket-molecule pairs. Without RAID acceleration, the same task required more than six days. Supplementary Methods and Supplementary Fig. 2(a-d) provide details of the first stage.

In the second stage, top-ranked molecules underwent post-processing refinement, which users may tailor to their preferred methods. A representative strategy involves clustering the top-ranked molecules and selecting diverse representatives (Fig. 3c,

right). These candidates were then refined by molecular docking (here, Glide-SP) for re-ranking and binding-pose prediction. For proteins with many available structures, all corresponding pockets or part of them were selected for docking. Results from independent docking runs were integrated using either intersection or union strategies, enhancing robustness against pocket conformational variability. Supplementary Methods and Supplementary Fig. 2(e-g) describe the details of the second stage.

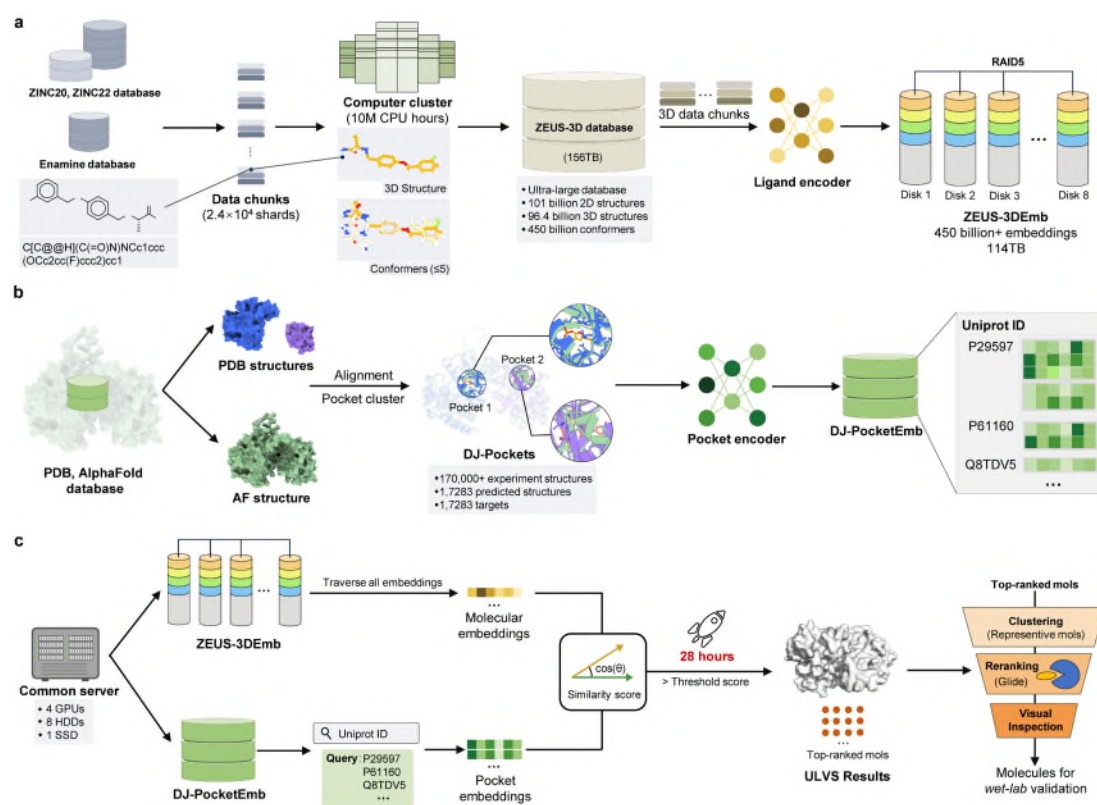


Fig. 3. Embedding generation and screening protocol. (a) Data processing of ZEUS-3D. Molecules from ZINC20, ZINC22, and Enamine Real were divided into over 24,000 shards, which were then allocated to 26 chunks. A computer cluster generated up to five 3D structures per molecule. ZEUS-3D has a storage capacity of 156 TB, containing 101 billion 2D molecules, 96.4 billion structures, and 450 billion 3D conformations. The molecule encoder of DrugJEPA computed embeddings for each conformation, yielding ZEUS-3DEmb with a storage size of 114 TB. RAID5 was used to enhance data reliability and I/O performance. (b) Construction of DJ-Pockets and DJ-PocketEmb. All PDB database targets were processed to build DJ-Pocket, a Uniprot ID-indexed pocket database for easy accessibility. For each target, all PDB structures were aligned to its AlphaFold structure, and pockets were extracted using 3D information of known PDB ligands. DJ-PocketEmb, the embedding counterpart of DJ-Pocket, was built by computing pocket embeddings using the pocket encoder of DrugJEPA. (c) ZEUS-3D screening protocol. Pocket embeddings were retrieved from DJ-PocketEmb using target Uniprot IDs and then used to scan ZEUS-3DEmb's molecule embeddings via cosine similarity. Molecules were retained if the similarity exceeded a threshold of 0.6 (orange dots). Traversal of ZEUS-3D was completed in 28

hours on a conventional server. Retained molecules underwent clustering; cluster representatives were reranked via docking (Glide-SP). Finally, docking results were visually inspected, and candidates were selected for wet-lab validation.

Wet-lab validation of DrugJEPa

To experimentally validate DrugJEPa, we evaluated its performance on three distinct protein targets: tyrosine kinase 2 (TYK2), trace amine-associated receptor 1 (TAAR1), and pseudouridylate synthase 1 homolog (PUS1), representing the kinase, G protein-coupled receptor (GPCR), and epigenetic target classes, respectively. Notably, there are currently no publicly reported small molecule ligands for PUS1 as far as we know, providing a particularly challenging test for the screening model.

Discovery of TYK2 inhibitors

TYK2 is a member of the Janus kinase (JAK) family and mediates signaling downstream of cytokine receptors, including those for IL-12, IL-23, and type I interferons. It plays a critical role in immune regulation and the pathogenesis of autoimmune diseases [51]. Consequently, TYK2 has emerged as a promising therapeutic target for various immune disorders and has garnered considerable attention in recent years. TYK2 contains multiple functional domains, with the C-terminal JH1 kinase domain and the JH2 pseudokinase domain together forming its core regulatory module. Selective targeting of the JH2 domain enables more precise modulation of TYK2 activity while minimizing off-target effects on other JAK family members [52-54]. Here, we selected the TYK2-JH2 domain as one of our target examples.

Using TYK2-JH2 structures (Supplementary Table 8), we performed an ultra-large virtual screen (ULVS) against the ZEUS-3D compound library. Top-ranked molecules from the screen were clustered, and representative compounds were selected to construct a pool of 10,000 potential hits. From this pool, we virtually selected 40 molecules that are relatively amenable to either chemical synthesis or commercial purchase with lower price (Supplementary Fig. 3) for further experimental evaluation. Among these, 16 compounds exhibited >50% inhibition of TYK2-JH2 at 10 μ M in the HTRF assay (Fig. 4a), corresponding to a hit rate of 40%. We next determined the half-maximal inhibitory concentration (IC_{50}) values of these active compounds (Fig. 4b and Supplementary Fig. 4). Six hits exhibited IC_{50} values below 1 μ M (Fig. 4(b-c)): TYK2-118 (1.6 nM), TYK2-023 (24.0 nM), TYK2-119 (44.0 nM), TYK2-102 (49.1 nM), TYK2-107 (0.36 μ M), and TYK2-26X (0.71 μ M). The interactions between these compounds and TYK2-JH2 were further validated by the differential scanning fluorimetry (DSF) assay (Fig. 4d and Supplementary Fig. 5). Among them, TYK2-118 induced the largest thermal shift, increasing the melting temperature by 16.38 $^{\circ}$ C at 20 μ M (Fig. 4d). Detailed information for all compounds is provided in Supplementary Fig. 3.

Finally, we determined the co-crystal structure of the TYK2-JH2 domain in complex with one of the most active compounds, TYK2-119 (Supplementary Table 9), at a resolution of 1.72 $\text{\AA}$ (PDB ID: 24HM). TYK2-119 occupies the classical binding pocket of TYK2-JH2 (Fig. 4e) and is well supported by the electron density map (Fig.

4f). TYK2-119 is primarily anchored through a hydrogen-bonding network: the carbonyl group of the lactam moiety forms a hydrogen bond with the amino group of Lys642, whereas the aminopyrimidine moiety forms two additional hydrogen bonds with the backbone NH and carbonyl groups of the hinge residue Val690. In addition, the hydrophilic tail, comprising a piperidin-4-ol group, extends toward the solvent-exposed region (Fig. 4g). Collectively, these interactions provide a structural rationale for the high inhibitory potency of TYK2-119.

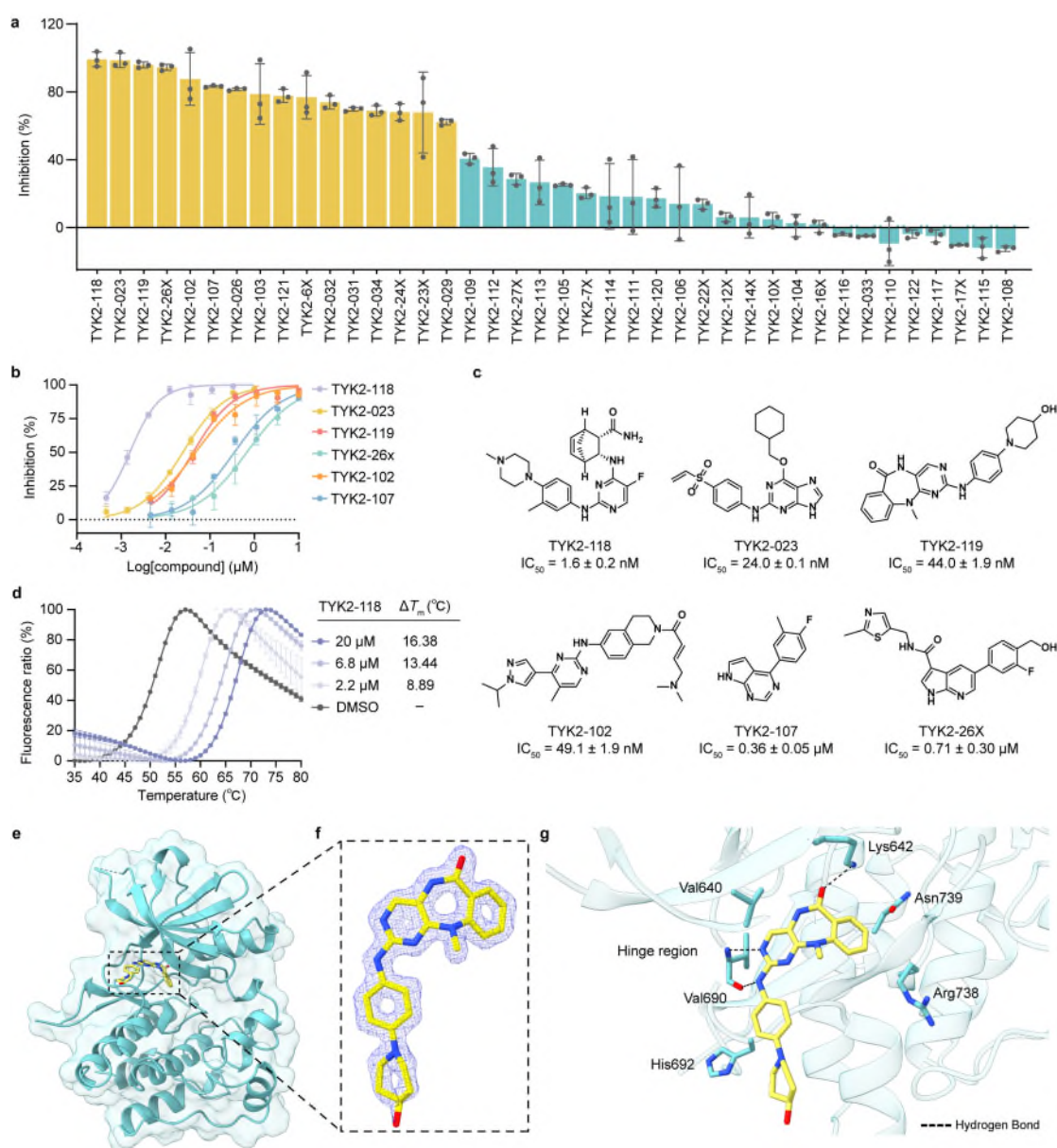


Fig. 4. Results of wet-lab validation of ULVS targeting TYK2-JH2. (a) Inhibition rate of each compound at 10 μM in the HTRF assay ($n = 3$). Compounds exhibiting >50% inhibition are highlighted in yellow. (b) Dose–response curves of the six most active compounds against TYK2-JH2 in the HTRF assay ($n = 3$). (c) Chemical structures of the six most active compounds. (d) Stabilizing effect of TYK2-118 on the TYK2-JH2 measured by the DSF assay ($n = 3$). (e) Overall structure of TYK2-JH2 in complex with

TYK2-119 (PDB ID: 24HM). TYK2-JH2 is shown as a cyan cartoon with a transparent surface, and TYK2-119 is shown as yellow sticks. (f) Electron density map of TYK2-119 in the binding pocket. (g) Binding mode of TYK2-119 with TYK2-JH2. Hydrogen-bonding interactions are shown as black dashed lines, and key interacting residues are labelled. The structure figures were prepared with ChimeraX. All data were obtained from three independent experiments. Data are shown as averages $\pm$ SD.

Discovery of TAAR1 Agonists

TAAR1 is a member of the trace amine-associated receptor (TAAR) family and belongs to class A GPCRs. As a key modulator of dopaminergic, serotonergic, and glutamatergic signaling in the brain, TAAR1 is implicated in the pathophysiology of several psychiatric disorders, including schizophrenia, bipolar disorder, major depressive disorder, and attention deficit hyperactivity disorder (ADHD) [55, 56]. Importantly, genetic polymorphisms in the TAAR1 gene have been linked to increased susceptibility to schizophrenia and substance use disorders—such as methamphetamine and cocaine addiction—due to the receptor's role in regulating reward pathways [55]. Given its strong disease relevance, TAAR1 has emerged as a promising therapeutic target for mental disorders.

As described above, we performed ULVS against the ZEUS-3D library and compiled a pool of 10,000 potential TAAR1 modulator hits taken from clustered top-ranked molecules from the screen. From this pool, we selected 40 molecules for chemical synthesis or commercial purchase, followed by experimental testing (Supplementary Fig. 6). Among these, eight compounds exhibited over 50% agonist activity at 100 μ M in the GloSensor cAMP assay—a relatively high screening concentration chosen because this class of GPCR targets is inherently more difficult to target effectively than kinases—resulting in a hit rate of 20.0% (Fig. 5a). We further determined the half-maximal effective concentration (EC_{50}) values for these eight compounds, six of which showed EC_{50} values below 50 μ M (Fig. 5(b-c); Supplementary Fig. 6; Supplementary Fig. 7). This activity was further confirmed using the HTRF assay, which yielded an EC_{50} of 97.9 nM (Fig. 5d). Notably, the most potent molecule, TAAR1-119, is asenapine—an approved antipsychotic drug known to target the 5-HT_{2A} receptor—whose TAAR1 activity was independently identified through our structure-based virtual screening pipeline. A subsequent literature review confirmed that this molecule has recently been reported to have off-target activity against TAAR1 [57]. Detailed information for all compounds is provided in Supplementary Fig. 6.

To understand the binding mode of the most active compound, TAAR1-119, with TAAR1, we attempted to determine the three-dimensional structure of the complex using cryo-EM or X-ray crystallography, but these efforts were unsuccessful. Therefore, we performed molecular docking (Glide XP) to predict their binding mode. As shown in Fig. 5e, the predicted binding pose suggests that TAAR1-119 is accommodated within the orthosteric binding pocket of TAAR1, forming π - π interactions with F268^{6.52} and F186^{ECL2}, a π -cation interaction with F267^{6.51}, along with hydrophobic interactions involving I104^{3.33}, V184^{ECL2}, and W264^{6.48}. To validate this predicted binding mode, we conducted site-directed mutagenesis experiments. Based on the predicted binding

mode, we selected six residues predicted to be important and introduced alanine mutations: F268^{6.52}A, F186^{ECL2}A, F267^{6.51}A, I104^{3.33}A, V184^{ECL2}A, and W264^{6.48}A. Our results showed that all these mutations markedly reduced TAAR1-119-induced cAMP signaling (Fig. 5f). Notably, the F267^{6.51}A, F186^{ECL2}A, and I104^{3.33}A mutants resulted in approximately 100-fold decreases in activity, indicating that these residues play critical roles in ligand recognition and stabilization.

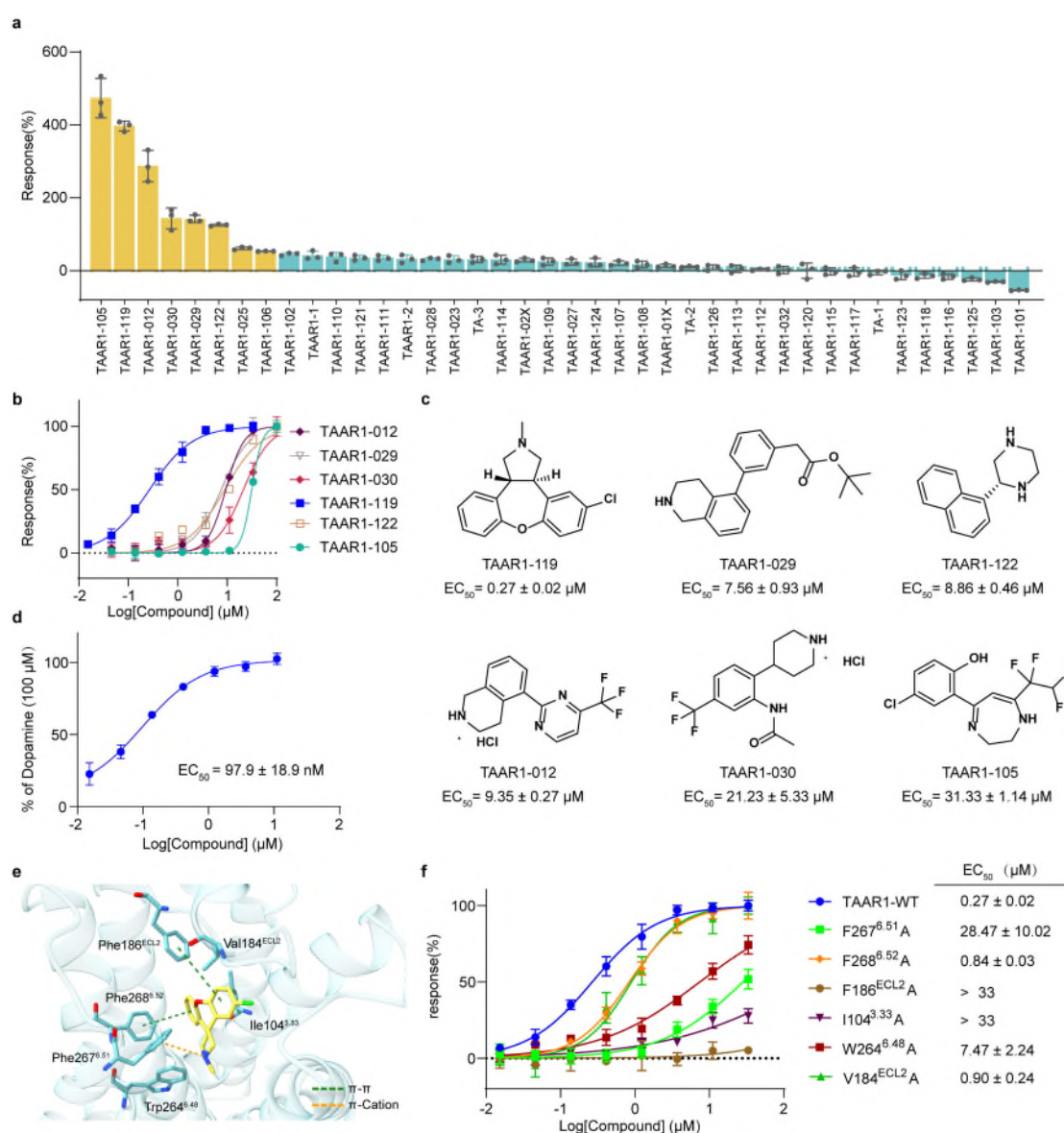


Fig. 5. Results of wet-lab validation of ULVS targeting TAAR1. (a) Inhibition rates of 40 selected compounds at 100 μ M in the GloSensor cAMP assay ($n = 3$). Compounds exhibiting >50% inhibition are highlighted in yellow. (b) Dose–response curves of the six most active compounds against TAAR1 in the GloSensor cAMP assay ($n = 3$). (c) Chemical structures of the six most active compounds. (d) Agonist activity of TAAR1-119 against TAAR1 measured by the HTRF assay ($n = 3$). (e) Predicted complex structure of TAAR1-119 bound to TAAR1 (PDB ID: 8JLO) using molecular docking

(Glide XP). The figure was prepared with ChimeraX. (f) Mutational effects of binding pocket residues on TAAR1-119-induced receptor activation, assessed by the GloSensor cAMP assay ($n = 3$). All data were obtained from three independent experiments. Data are shown as averages $\pm$ SD.

Discovery of PUS1 inhibitors

PUS1 is an RNA-modifying enzyme that catalyzes the conversion of uridine to pseudouridine (Ψ), the most abundant RNA modification, which enhances mRNA stability and translation [58, 59]. Numerous studies indicate that PUS1 not only serves as a critical regulator of RNA biology but also represents a promising potential target for the treatment of cancer and other diseases associated with dysregulated Ψ modification [60-63]. Consequently, developing small-molecule inhibitors of PUS1 may offer a novel therapeutic strategy for these conditions. Unlike TYK2 and TAAR1, for which many known ligands have been reported, no small-molecule inhibitors of PUS1 have been described to date, making PUS1 a challenging protein target.

We again conducted ULVS for PUS1 inhibitors against the ZEUS-3D library. Top-ranked molecules from the screening were clustered, and representative compounds were selected to construct a pool of 10,000 potential hits. From this pool, we virtually selected 40 molecules that are easier to synthesize or purchase for further experimental evaluation (Supplementary Fig. 8). To assess the bioactivity of compounds against PUS1, we established a fluorescence polarization (FP)-based assay using FAM-labeled mRNA. Inhibition rate assays demonstrated that 12 out of the 40 molecules exhibited significant PUS1 inhibitory activity (inhibition rate $> 50\%$ at $50 \mu\text{M}$) (Fig. 6a), resulting in a hit rate of 30%. We further determined the IC_{50} values of these 12 active molecules, among which 5 molecules showed IC_{50} values lower than $10 \mu\text{M}$ (Fig. 6(b-c); Supplementary Fig. 9). To further substantiate the bioactivity of compounds, we employed the DSF assay, which demonstrated that these five compounds were able to substantially stabilize the PUS1 protein (Supplementary Fig. 10). Notably, the most active compound, PUS1-110, achieved a sub-micromolar IC_{50} value of $0.85 \mu\text{M}$ in the FP assay (Fig. 6(b-c)). It induced a thermal shift (ΔT_m) value of 3.56°C at a concentration of $66.67 \mu\text{M}$ in the DSF assay (Supplementary Fig. 10) and exhibited a dissociation constant (K_D) value of $3.91 \mu\text{M}$ in the isothermal titration calorimetry (ITC) assay (Fig. 6d).

We then attempted to resolve the co-crystal structure of PUS1 in complex with PUS1-110, but these efforts were also unsuccessful. We therefore performed molecular docking (Glide XP) to predict the binding mode of PUS1-110 with PUS1. The docking results suggested that PUS1-110 fits perfectly within the catalytic pocket of PUS1, thereby blocking PUS1 from binding to RNA (Fig. 6e). PUS1-110 is a symmetric molecule consisting of two same substituted naphthalene rings linked by a urea bond, and it forms multiple interactions within the catalytic site of PUS1. Specifically, the carbonyl group of the urea moiety forms a hydrogen bond with Lys193. One naphthalene ring (right side, Fig. 6e) engages in π -alkyl interactions with the backbone carbon of Arg144, while its hydroxyl substituent forms a hydrogen bond with Tyr201. One oxygen atom of the sulfonic acid substituent participates in an electrostatic

interaction with Arg295, and another oxygen atom forms a hydrogen bond with the backbone amino group of His292. The other naphthalene ring (left side, Fig. 6e) forms a π -cation interaction with Arg261. The hydroxyl substituent on this naphthalene ring forms a hydrogen bond with Arg199, and one oxygen atom of the sulfonic acid substituent forms an electrostatic interaction with Arg261.

To validate this predicted binding mode, we performed site-directed mutagenesis experiments. Based on the docking results, PUS1-110 was predicted to establish critical interactions with seven residues of PUS1: R144, K193, R199, Y201, R261, H292, and R295. Each of these residues was individually mutated to alanine, generating seven mutants (R144A, K193A, R199A, Y201A, R261A, H292A, and R295A). We then assessed the bioactivity of PUS1-110 against these mutants using DSF assays. Our results showed that six mutations—R144A, K193A, Y201A, R261A, H292A, and R295A—significantly reduced the bioactivity of PUS1-110, thereby validating the predicted binding mode of PUS1-110 with PUS1 (Fig. 6f).

To the best of our knowledge, this study represents the first report of small-molecule inhibitors of PUS1. This finding not only fully validates the reliability and utility of the model and ultra-large-scale virtual screening strategy developed in this study but also provides the first highly potent lead compound for the field of PUS1-targeted small-molecule inhibitor research, laying a solid foundation for subsequent drug optimization and mechanistic studies.

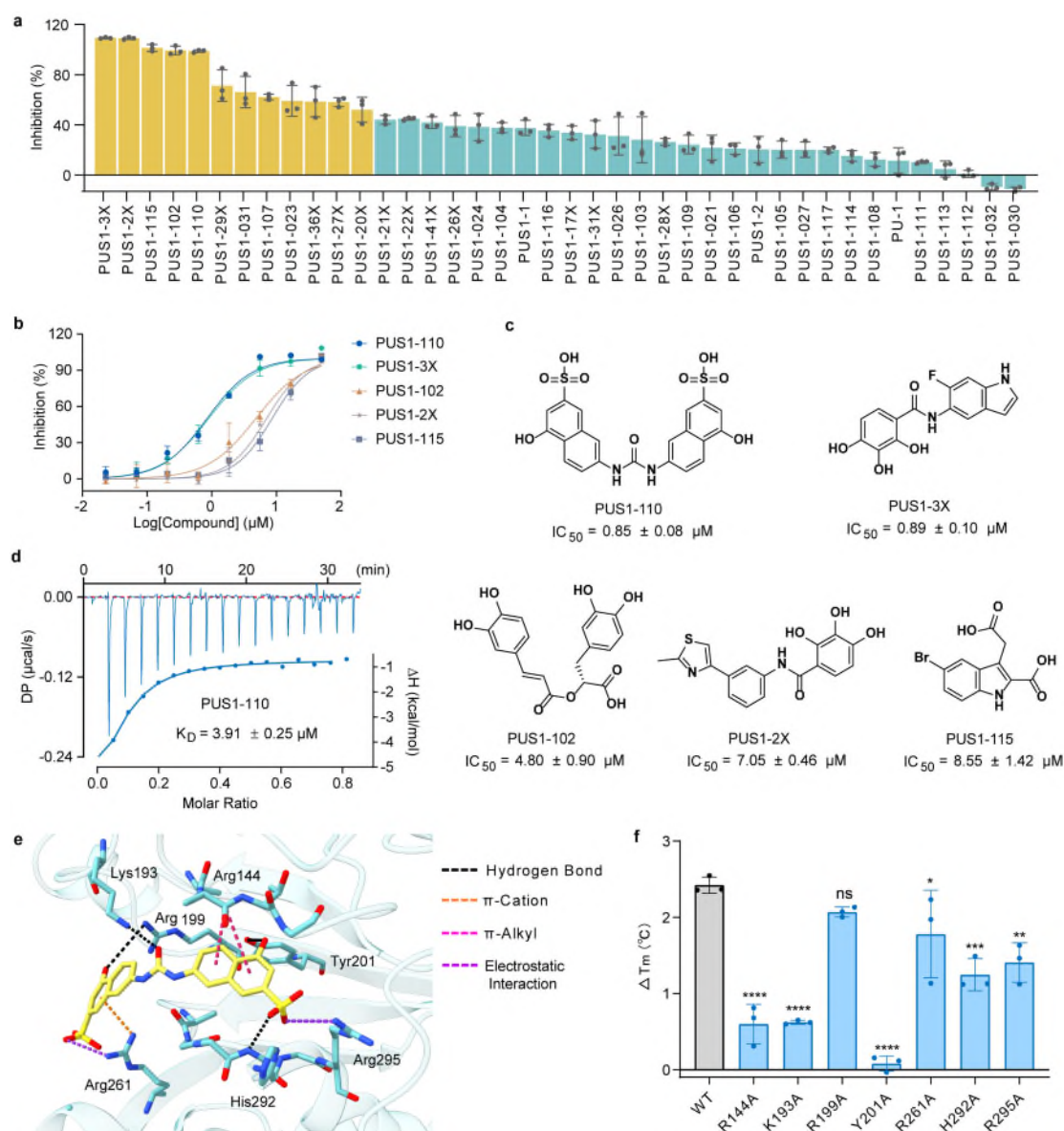


Fig. 6. Results of wet lab validation of ULVS targeting PUS1. All the experiments were performed independently three times. (a) Inhibition activity of each compound at $50\mu\text{M}$. (b) Dose-activity curves of the five most active hits against PUS1 in the FP assay. (c) Structures of the five most active hits. (d) Results from the ITC assay for PUS1-110. (e) A predicted binding mode of PUS1-110 with PUS1 using Glide-XP (the 3D structure of PUS1 was taken from PDB ID: 4nz7). Figures were prepared using ChimeraX. (f) DSF analysis of PUS1-110 ($22.2 \mu\text{M}$) binding to PUS1 WT and mutant proteins. Statistical analyses were performed by one-way ANOVA with Dunnett's multiple comparisons test. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$)

Discussion

Chemical libraries play a critical role in modern drug discovery, providing a source for retrieval of initial active compounds. Nevertheless, the relative advantages of large-scale versus small-scale databases in virtual screening continue to be debated. Central to this discussion are the trade-offs between computational cost and false positive rates.

Screening against large-scale databases generally demands greater computational resources and tends to produce a higher number of false positives, which in turn necessitates expensive downstream experimental validation [3, 9, 64]. Conversely, small-scale databases are computationally less intensive but may suffer from sample bias, potentially leading to missed opportunities in identifying truly active compounds. Ideally, researchers would be able to screen large-scale databases both affordably and accurately, as doing so increases the likelihood of discovering high-quality active compounds.

Traditionally, the pharmaceutical industry has relied on two main types of physical compound collections: proprietary in-house libraries and commercially available in-stock collections. However, these physical libraries face inherent limitations—such as high handling costs, slow linear growth, finite size, and diminishing novelty—which increasingly constrain their ability to explore expansive chemical diversity. A transformative shift is now underway with the emergence of virtual on-demand chemical spaces. These platforms enable the rapid, parallel synthesis of compounds from available building blocks using validated chemical protocols. Crucially, they ensure significant chemical novelty and support polynomial library growth by seamlessly integrating new synthons and advanced reaction scaffolds, including multi-component reactions. Notable examples such as ZINC22 [2] and Enamine [45] now provide access to compound collections on the scale of 100 billion. Another style is represented by exhaustive brute-force enumeration, focusing on the exploration of novel chemical scaffolds beyond existing combinatorial spaces, exemplified by GDB-17 [65] and GDB-20 [66]. As the scale of modern chemical space increasingly exceeds the capacity of traditional screening methods, there is a clear need for advanced tools capable of navigating vast datasets while balancing speed, accuracy, and cost. To address this challenge, we introduce DrugJEPA—a deep learning framework that integrates contrastive learning with the JEPA and MoE approaches.

Leveraging several novel algorithmic innovations, DrugJEPA achieves state-of-the-art (SOTA) performance. Most existing contrastive learning (CL)-based virtual screening (VS) methods optimize InfoNCE-style objectives to maximize the lower bound of cross-modal mutual information [20, 22]. While effective for global discriminative alignment, these objectives do not explicitly constrain the uncertainty of cross-modal representations in the joint embedding space. Consequently, embeddings trained solely on InfoNCE loss lack fine-grained predictability, limiting their capacity to capture subtle ligand-protein interaction patterns [67]. In this work, we propose DrugJEPA, which integrates JEPA—an energy-based model paradigm [29]—into contrastive learning for VS. By minimizing a non-negative energy function, the model assigns lower energy to positive ligand-pocket pairs than to negative ones, aligning naturally with the physical intuition of molecular recognition. Beyond discrimination, DrugJEPA introduces a latent predictor to project input-modality embeddings into the target embedding space, explicitly enforcing cross-modal predictability. Thermodynamically, InfoNCE-based learning primarily boosts mutual information [68, 69] (analogous to lowering effective internal energy during ligand binding) while leaving cross-modal conditional uncertainty under-constrained; DrugJEPA’s complementary L2 loss addresses this by reducing conditional entropy, thereby enhancing predictability and stabilizing the joint embedding structure [70]. To better

reflect biological reality, we adopt a multi-label alignment strategy that aligns each protein with multiple ligands simultaneously, and introduce a rank loss to model relative binding affinities, enabling both binder/non-binder discrimination and ligand potency ranking. We further integrate protein sequence-derived evolutionary information and a Mixture-of-Experts (MoE) architecture for fine-grained feature extraction. Comprehensive benchmark evaluations confirm that DrugJEPA consistently outperforms classical docking tools (Glide, Vina) and establishes a new SOTA for deep learning-based VS.

In addition to its outstanding screening performance, we have built an ULVS platform based on DrugJEPA, which enables ultrafast virtual screening on affordable hardware. Unlike traditional molecular docking, which requires iterative energy minimization and scoring for every ligand-protein pair, DrugJEPA transforms the full VS pipeline into one-time embedding computation followed by high-throughput vector retrieval, delivering up to seven orders of magnitude speedup over docking and enabling computationally feasible billions-scale screening. Scaling to 100 billion molecules brings critical systems engineering challenges, mainly massive embedding dataset storage and I/O bottlenecks, as I/O operations (not GPU computation) become the dominant performance limiter for ultra-large datasets. To overcome this, we implemented three optimizations: converting embeddings to a compact sequential-access binary format to cut storage footprint and I/O latency; deploying a RAID 5 HDD configuration to balance cost, read performance, and data redundancy; and an asynchronous parallel scheduling framework overlapping data loading and GPU inference (details see Supplementary Methods). These strategies significantly reduce I/O latency, improve overall system efficiency, and deliver these performance gains with low hardware expenditure.

Despite the evident utility of DrugJEPA in ultra-large-scale virtual screening, several limitations remain. A key drawback is its limited interpretability: unlike traditional or advanced AI-based docking methods, DrugJEPA does not predict binding poses, thereby restricting mechanistic insights into ligand-pocket interactions. Future work will focus on enhancing interpretability to generate testable binding hypotheses. Another issue stems from the ultra-large scale of the database, which often results in the retrieval of a vast number of potential active molecules, making it difficult to prioritize candidates for experimental activity testing. In this study, we first performed random sampling of the primary screening hits to control the compound set size while maximizing structural diversity and chemical space coverage, followed by molecular clustering to generate representative molecules and reduce the compound number. These representative compounds were then subjected to independent molecular docking against multiple binding pockets of the target. The top-ranked molecules were selected from each docking result, counted the occurrence frequency of each molecule, and chose those that recurred across multiple docking tasks, thereby determining the final candidate compounds for wet-lab experiments. This strategy considers both screening speed and the ability to identify molecules with novel scaffold structures. Other approaches for selecting candidate molecules, such as combining with agents, warrant further exploration. Despite these limitations, our model effectively balances screening speed, accuracy, and economic cost, offering an affordable ultrafast,

accurate virtual screening system for ultra-large-scale chemical libraries.

Methods

Model Evaluation

The DUD-E dataset contains 22,886 active compounds against 102 targets, and 50 decoys for each active. The LIT-PCBA has 15 targets, 7761 actives and 382674 decoys. Compared with DUD-E, LIT-PCBA is a more challenging benchmark for virtual screening, as it employs experimentally confirmed actives and inactives from PubChem [71] HTS (High Throughput Screening), providing an unbiased benchmark that reflects real-world virtual-screening difficulty. All structures of targets in the two datasets have been removed from the training set for ensuring fair comparison. We choose three classic metrics to evaluate the power of virtual screening: AUROC (Area Under the Receiver Operating characteristic Curve), BEDROC (Boltzmann-Enhanced Discrimination of ROC) [72] and EF (Enrichment Factor). AU-ROC measures the discriminatory power of a model in ranking active ligands ahead of inactive counterparts, defined as the integral under the curve relating true positive rate to false positive rate. BEDROC extends AU-ROC by introducing exponential weighting, thereby placing greater emphasis on the early recognition of active compounds. The EF-1% is a widely adopted metric in virtual screening studies, which quantifies ability of model to identify actives among top 1% ligands. We compared DrugJEPA with multiple baseline methods, including traditional docking methods [73, 74], classic machine learning-based methods [75-77], contrastive learning-based methods [20-22, 41, 78]

Mixture of Experts (MoE) Block

The mixture-of-experts block is primarily composed of two key components: the sparse MoE layer and the expert router. Similar to ensemble learning, the sparse MoE consists of a set of E individual networks (typically multilayer perceptrons, MLPs). Each network in this set, referred to as an "expert," is responsible for processing a different subset of the training samples, focusing on a specific region of the input space. The expert router is used to select the appropriate experts for processing a given input; it determines the weights assigned to each expert and selectively activates the top- K experts for each input token. During training, both the experts and the gating networks are trained simultaneously to optimize their performance and decision-making ability. In Transformers, the MoE block can replace the feed-forward network (FFN) to achieve improved model performance.

Specifically, a router network selects Top- K experts out of S total experts to scale up the model with multiple MLP blocks in parallel. This router network is often a linear layer that computes the normalized weight matrix based on the inputs $\mathbf{X} \in \mathbb{R}^{N \times D}$ for voting, where N is the number of tokens and D is the dimension of input, resulting in

$$\mathbf{W}_E = \text{Softmax}(\text{Lin}(\mathbf{X})) \quad (1)$$

The Top- K experts are selected for each token based on $\mathbf{W}_E \in \mathbb{R}^{N \times S}$, and the re-

normalized weights $\mathbf{W}_K \in \mathbb{R}^{N \times K}$ are computed using:

$$\mathbf{W}_K = \text{Softmax}(\text{TopK}(\mathbf{W}_E)) \quad (2)$$

Each selected expert is represented by an MLP block, and the output is obtained through a re-weighted sum and residual connection:

$$\mathbf{X}_{\text{out}} = \mathbf{X} + \sum_{e \in \text{Top-K}} \mathbf{W}_K^e \cdot \text{MLP}_e(\mathbf{X}) \quad (3)$$

Co-Upcycling of MoE blocks

Training transformers with sparse MoE blocks from scratch can often be unstable and expensive. To mitigate these challenges, we adopt a co-upcycling approach. Specifically, for each sparse transformer layer, the attention block is initialized using weights from a pretrained dense transformer layer (Unimol was used in this work). In the MoE block, expert layers are derived from multiple copies of the FFN block in the pretrained dense transformer, while the router network is randomly initialized [35]. This strategy consistently enhances training stability and overall model performance. We adopted alternating architecture, where every other transformer block is converted to MoE, yielding a [sparse, dense] rhythm [37].

Training

For model training, the pocket and ligand encoders were initialized with UniMol [79] pre-trained weights, and converted to sparse Mixture of Experts (MoE) [32] encoders via the Co-upcycling strategy [35]. The protein sequence encoder adopted the ESMC-600M model [38] as the backbone; all sequence embeddings were precomputed and stored locally to avoid redundant computation during training. A differential parameter update strategy was applied: full parameters of the pocket and ligand encoders were updated, while the ESMC backbone was frozen with only the sequence mapper parameters updated. The model was trained for 60 epochs using the Adam optimizer [80] (learning rate = 1×10^{-4}) on 4 NVIDIA H200 GPUs.

Training Data

The dataset used for DrugJEPA was constructed with the systematic curation pipeline established by Feng et al. [21], which was developed to generate high-quality pocket-ligand binding affinity data sourced from ChEMBL [81] and BindingDB [82, 83]. First, affinity data were organized into assays to ensure comparability under consistent experimental conditions, followed by assay-guided pocket matching using protein sequence similarity and ligand Tanimoto similarity, with irrelevant pockets filtered via BioLip2 [84]. To enhance protein diversity and improve model generalization, additional protein-ligand complexes from PDDBind [85, 86] were incorporated, yielding a final dataset covering 4,847 proteins and 53,406 pockets.

Training Loss

Contrastive Loss. The goal of the contrastive loss is to maximize the similarity between active pocket-ligand pairs, while simultaneously minimizing the similarity of non-active pairs. Based on this principle, we adopt a multi-sample aligned contrastive learning strategy, where the embedding of a single pocket is aligned with the embeddings of multiple active ligands. This strategy reflects an essential biological fact: a protein pocket is generally capable of binding to diverse ligand molecules.

Formally, within a training batch, let B_p and B_l denote the set of ligands and pockets sampled in this training batch, respectively. For a given pocket embedding $\mathbf{p}_i$ and its associated set of active ligands D_i , the contrastive objective consists of two directional terms: a pocket-to-ligand loss $\mathcal{L}_i^{p \rightarrow l}$, and a ligand-to-pocket loss $\mathcal{L}_i^{l \rightarrow p}$.

$$\mathcal{L}_i^{p \rightarrow l} = -\frac{1}{|D_i|} \sum_{k \in S_i} \log \frac{\exp\left(\frac{\text{Cos}(\mathbf{p}_i, \mathbf{l}_k)}{\theta}\right)}{\sum_{m \in B_l} \exp\left(\frac{\text{Cos}(\mathbf{p}_i, \mathbf{l}_m)}{\theta}\right)} \quad (4)$$

$$\mathcal{L}_i^{l \rightarrow p} = -\frac{1}{|D_i|} \sum_{k \in S_i} \log \frac{\exp\left(\frac{\text{Cos}(\mathbf{p}_i, \mathbf{l}_k)}{\theta}\right)}{\sum_{n \in B_p} \exp\left(\frac{\text{Cos}(\mathbf{p}_n, \mathbf{l}_k)}{\theta}\right)} \quad (5)$$

Similar to the contrastive losses described above, we further introduce a contrastive objective between sequence and ligand $\mathcal{L}_i^{l \rightarrow seq}$ and $\mathcal{L}_i^{seq \rightarrow l}$. The motivation is to explicitly align ligand representations not only with their corresponding binding pockets, but also with protein sequences of the same targets. In this way, ligand embeddings are jointly constrained by both structural and sequential protein information, leading to a more robust and biologically consistent representation space. The final contrastive learning objective combines both modalities and is expressed as:

$$\mathcal{L}_{contrast} = \frac{1}{2} \sum_{i=1}^{B_p} (\mathcal{L}_i^{p \rightarrow l} + \mathcal{L}_i^{l \rightarrow p}) + \frac{1}{2} \sum_{i=1}^{B_p} (\mathcal{L}_i^{l \rightarrow seq} + \mathcal{L}_i^{seq \rightarrow l}) \quad (6)$$

Regularization Loss. The regularization loss measures the squared L2 distance between two modalities at latent representation space; minimizing it drives the encoder of one modality (e.g., ligands) to match the latent vectors of the other (e.g., pockets), thus training the model to minimize the uncertainty (conditional entropy) of representations, thereby learning deterministic and predictable latent patterns. The regularization loss of DrugJEPA is defined as L2 distance:

$$\mathcal{L}_{reg} = \frac{(\mathcal{L}_{lig \rightarrow pkt} + \mathcal{L}_{pkt \rightarrow lig})}{2} \quad (7)$$

$$\mathcal{L}_{lig \rightarrow pkt} = \|f_{lig \rightarrow pkt}(\mathbf{l}) - \mathbf{p}\|_2^2 \quad (8)$$

$$\mathcal{L}_{pkt \rightarrow lig} = \|f_{pkt \rightarrow lig}(\mathbf{p}) - \mathbf{l}\|_2^2 \quad (9)$$

where $f_{lig \rightarrow pkt}$ denotes an MLP that predicts the pocket embedding from the ligand embedding $\mathbf{l}$, whereas $f_{pkt \rightarrow lig}$ predicts the ligand embedding from the pocket embedding $\mathbf{p}$.

Ranking Loss. For each assay A_i , ligands are ranked according to their experimentally measured affinities, producing an ordered sequence:

$$s = [l_i^1, l_i^2, \dots, l_i^B] \quad (10)$$

The superscripts are the ranked index in the ordered sequence. The subscript i denotes the index of assay. In line with the approaches of Cao et al. [30] and Feng et al. [14], the likelihood of choosing ligand l_i^k at the k -th step (from the remaining candidate set θ_i^k from k -th to B -th ligands) is defined as:

$$p_i^k = \frac{e^{c_i^k}}{\sum_{j \in \theta} e^{c_i^j}} \quad (11)$$

where c_i^j is the cosine similarity between ligand and pocket (or sequence) embedding.

The rank loss for assay A_i is defined as:

$$\mathcal{L}_{\text{rank}} = - \sum_{k=1}^B \frac{1}{\sqrt{B} \log(k+1)} \log(p_i^k) \quad (12)$$

Auxiliary Loss for MoE. In this work, the auxiliary loss of MoE has two parts: load balancing loss and router z-loss. The load balancing loss is a crucial mechanism for addressing routing collapse issues. Its core objective is to ensure balanced utilization of all experts, preventing a scenario where a few experts are excessively activated while others remain underutilized. Through optimizing load balancing loss, MoE model encourage experts to receive roughly equal numbers of training examples. Given E experts, the load balancing loss is defined as following:

$$\mathcal{L}_{\text{Balance}} = \alpha \cdot E \cdot \sum_{e=1}^E f_e \cdot P_e \quad (13)$$

where α is loss weight coefficient (hyperparameter, controlling regularization intensity), whose value is 0.01 in this work; f_e denotes the ratio of the number of tokens that are dispatched to e -th expert to the total number of tokens in the current batch.

$$P_e = \frac{1}{T} \sum_{i=1}^T p_{e,i} \quad (14)$$

where $p_{e,i}$ is the logit assigned by the routing network to expert e for token i ; T is the total number of tokens in current batch. When the values of both f_e and P_e for all experts are close to E , the loss is minimized, resulting in the most balanced utilization of the experts.

In MoE models, router z-loss helps to prevent excessively large logits that could destabilize the gradient during training, ensuring robust and balanced model performance. Router z-loss directly penalizes the magnitude of the logits, thereby

constraining their numerical range:

$$\mathcal{L}_z = \beta \cdot \frac{1}{T} \sum_{i=1}^T \left(\log \sum_{e=1}^E \exp(x_e^{(i)}) \right)^2 \quad (15)$$

where $x_e^{(i)}$ denotes the logits of the i -th token for the e -th expert; β is the loss weight coefficient, whose value is 0.001 in this work. During training, minimizing the output of the $\mathcal{L}_z$ force the logits to contract toward zero, thereby reducing the gradient sensitivity of the softmax function.

Above all, the total loss of MoE model is a sum of the contrastive loss ($\mathcal{L}_{Contrast}$), regularization loss ($\mathcal{L}_{reg}$), the auxiliary load balance loss ($\mathcal{L}_{Balance}$), and the router z-loss ($\mathcal{L}_z$), yielding a total loss:

$$\mathcal{L}_{total} = \mathcal{L}_{Contrast} + \mathcal{L}_{reg} + \mathcal{L}_{Balance} + \mathcal{L}_z \quad (16)$$

Differential scanning fluorimetry assay for TYK2

DSF experiments of TYK2 were performed on a Bio-Rad CFX Connect Real-Time System. Each 10 μ L reaction mixture contained 2 μ M recombinant TYK2 JH2 (residues 575–869), 100 μ M test compound, 5 $\times$ SYPRO Orange dye, and assay buffer (20 mM Tris-HCl, pH 7.7, 150 mM NaCl). Thermal denaturation was carried out by ramping the temperature from 25 $^{\circ}$ C to 95 $^{\circ}$ C at 1 $^{\circ}$ C/min.

DSF assays of PUS1 were conducted using a real-time PCR detection platform. SYPRO Orange (Sigma, 5000 $\times$ stock solution, cat. no. S5692) was employed as the fluorescent reporter dye. All DSF measurements were performed in a buffer system consisting of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. The excitation and emission wavelengths were configured at 492 nm and 610 nm, respectively. The reaction cocktail, containing 2 μ M PUS1 protein, 5 $\times$ SYPRO Orange, and the test compounds, was subjected to a thermal gradient ranging from 25 $^{\circ}$ C to 95 $^{\circ}$ C, with a heating rate of 1 $^{\circ}$ C per 15 seconds. The melting temperature (T_m), defined as the inflection point of the protein unfolding transition curve, was calculated by fitting the data to the Boltzmann equation. All data processing and quantitative analysis were implemented using GraphPad Prism software.

Homogeneous time resolved fluorescence assay for TYK2

HTRF signals were quantified using the KINASE-6HIS BINDING DISCOVERY KIT (62KBD02PEA, PerkinElmer). Recombinant His-tagged TYK2 JH2 (residues 575–869) was pre-incubated with serially diluted test compounds for 10 min, followed by addition of anti-6His-Eu cryptate antibody and Dasatinib-d2, and incubation for an additional 30 min prior to TR-FRET measurement.

Crystallization, data collection, and structure determination of TYK2

Crystals were grown by vapor diffusion at 18 $^{\circ}$ C. Recombinant TYK2 JH2 (residues 575–869) at 8 mg/mL was preincubated with compound at a 1:3 molar ratio. The protein–ligand complex was then mixed 1:1 (v/v) with crystallization buffer containing

0.2 M ammonium citrate tribasic (pH 7.0), 0.1 M imidazole (pH 7.0), and 20% (w/v) polyethylene glycol monomethyl ether 2000. Diffraction data were collected at 0.9791 Å on beamline BL19U1 at the Shanghai Synchrotron Radiation Facility. Data were processed using HKL-3000 or XDS. Structure was solved by molecular replacement in CCP4, using the coordinates from PDB entry 7K7Q as the search model. The inhibitor was modeled manually in Coot, followed by iterative refinement in Phenix. Data collection and refinement statistics are summarized in Supplementary Table 9.

Cell lines for TAAR1

Human embryonic kidney 293 cells (HEK293T) were obtained from Cell Bank of the Chinese Academy of Sciences (CBCAS, China) and cultured in DMEM supplemented with 10% (v/v) FBS, 100 units/mL penicillin plus 100 µg/mL streptomycin at 37 °C in 5% CO₂.

cAMP accumulation assay for TAAR1

The cAMP accumulation assay was performed using the GloSensor -cAMP assay and the HTRF-cAMP dynamic kit. The GloSensor cAMP assay was conducted per the manufacturer's protocol (Promega). Intracellular cAMP levels were measured using the HTRF-cAMP dynamic kit (Revvity) in 384-well low-volume plates (10,000 cells per well) according to the manufacturer's instructions. The assay was performed on a CLARIO Star microplate reader (BMG Labtech).

Protein expression and purification for PUS1

The human PUS1 protein (amino acid residues 79–408) was subcloned into the pGEX-6P-1 expression vector. The recombinant protein was overexpressed in *Escherichia coli* BL21 (DE3) competent cells cultured in Luria Broth (LB) medium supplemented with ampicillin. Protein expression was induced by the addition of 0.2 mM isopropyl-1-thio-D-galactopyranoside (IPTG), followed by incubation at 16 °C for 15 h. The GST fusion tag was cleaved using PreScission Protease. All recombinant proteins were further purified via size-exclusion chromatography on a SuperdexTM 200 Increase 10/300 GL column (GE Healthcare). This protein requires two consecutive size-exclusion chromatography runs to remove nucleic acid contamination. The purified proteins were aliquoted and stored at -80 °C in a buffer consisting of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. Protein purification was performed on an ÄKTA pure system (Cytiva), and data acquisition was conducted using UNICORN 7.0 software. All PUS1 mutant proteins were expressed and purified using the same protocol as the wild-type protein.

Fluorescence polarization (FP) assay for PUS1

FP assays were conducted in a buffer composed of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.00001 g/mL BSA. 300 nM PUS1 was incubated with 4 nM FAM-labeled mRNA probe (FAM-mRNA). The sequence of the FAM-mRNA probe is AAAUCGGGAUCCU-FAM [87]. The FAM moiety is a fluorescent dye whose polarization level fluctuates with its binding state, thus enabling the detection of intermolecular interactions. All assays were implemented in 384-well black microplates, and FP signals were recorded with a CLARIOstar PLUS microplate reader. For

compound screening, test compounds were pre-incubated with the target protein for 30 min, after which the fluorescent probe was added to the reaction system, followed by an additional 8 h of incubation. To reduce the risk of RNase contamination, RNase-free water was utilized for the preparation of mRNA solutions, and all FP experiments were performed with RNase-free consumables. The acquired experimental data were analyzed and fitted via GraphPad Prism software.

Compounds availability

Molecules were obtained either via chemical synthesis or commercial purchase. Commercially sourced compounds were assigned numerical identifiers starting from 100 and were all purchased from TargetMol (www.targetmol.cn). Synthetic molecules were numbered from 1. The synthetic routes and characterization data for the hits are detailed in the Supplementary Methods.

References

1. Sadybekov, A.V. and V. Katritch, *Computational approaches streamlining drug discovery*. Nature, 2023. **616**(7958): p. 673-685.
2. Tingle, B.I., et al., *ZINC-22—A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery*. Journal of Chemical Information and Modeling, 2023. **63**(4): p. 1166-1176.
3. Fang, Z.H., et al., *Hit identification in ultra large virtual screening: an integrative review and future challenges*. Drug Discovery Today, 2026. **31**(2): p. 104616.
4. Macalino, S.J.Y., et al., *Role of computer-aided drug design in modern drug discovery*. Archives of Pharmacal Research, 2015. **38**(9): p. 1686-1701.
5. Sabe, V.T., et al., *Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review*. European Journal of Medicinal Chemistry, 2021. **224**: p. 113705.
6. Chen, X., et al., *Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification*. Signal Transduction and Targeted Therapy, 2020. **5**(1): p. 72.
7. Wang, Y., et al., *Modeling protein–ligand interactions for drug discovery in the era of deep learning*. Chemical Society Reviews, 2025. **54**(23): p. 11141-11183.
8. DiMasi, J.A., H.G. Grabowski, and R.W. Hansen, *Innovation in the pharmaceutical industry: New estimates of R&D costs*. Journal of Health Economics, 2016. **47**: p. 20-33.
9. Liu, F., et al., *The impact of library size and scale of testing on virtual screening*. Nature Chemical Biology, 2025. **21**(7): p. 1039-1045.
10. Cabeza de Vaca, I., et al., *Ultra-large virtual screening unveils potent agonists of the neuromodulatory orphan receptor GPR139*. Nature Communications, 2025. **17**(1): p. 129.
11. Gorgulla, C., et al., *An open-source drug discovery platform enables ultra-large virtual screens*. Nature, 2020. **580**(7805): p. 663-668.
12. Liu, F., et al., *Large library docking identifies positive allosteric modulators of the calcium-sensing receptor*. Science, 2024. **385**(6715): p. eado1868.
13. Lyu, J., et al., *Ultra-large library docking for discovering new chemotypes*. Nature, 2019. **566**(7743): p. 224-229.

14. Cherkasov, A., *The 'Big Bang' of the chemical universe*. Nature Chemical Biology, 2023. **19**(6): p. 667-668.
15. Zhang, K., et al., *Artificial intelligence in drug development*. Nature Medicine, 2025. **31**(1): p. 45-59.
16. Sadybekov, A.A., et al., *Synthon-based ligand discovery in virtual libraries of over 11 billion compounds*. Nature, 2022. **601**(7893): p. 452-459.
17. Gorgulla, C., et al., *AI-Enhanced Adaptive Virtual Screening Platform Enabling Exploration of 69 Billion Molecules Discovers Structurally Validated FSP1 Inhibitors*. bioRxiv, 2025: p. 2023.04.25.537981.
18. Radford, A., et al. *Learning transferable visual models from natural language supervision*. in *International conference on machine learning*. 2021. PmLR.
19. Chen, T., et al. *A simple framework for contrastive learning of visual representations*. in *International conference on machine learning*. 2020. PmLR.
20. Jia, Y., et al., *Deep contrastive learning enables genome-wide virtual screening*. Science, 2026. **391**(6781): p. eads9530.
21. Feng, B., et al., *Hierarchical affinity landscape navigation through learning a shared pocket-ligand space*. Patterns, 2025. **6**(10).
22. Han, J., Y. Hong, and W.-J. Li. *DrugHash: Hashing Based Contrastive Learning for Virtual Screening*. 2025.
23. LeCun, Y., et al., *A tutorial on energy-based learning*. 2006. **1**(0).
24. Grathwohl, W., et al., *Your classifier is secretly an energy based model and you should treat it like one*. 2019.
25. LeCun, Y. and F.J. Huang. *Loss functions for discriminative training of energy-based models*. in *International workshop on artificial intelligence and statistics*. 2005. PMLR.
26. Song, Y. and D.P.J.a.p.a. Kingma, *How to train your energy-based models*. 2021.
27. Balestriero, R. and Y. LeCun, *LeJEPa: Provable and Scalable Self-Supervised Learning Without the Heuristics*. arXiv preprint arXiv:2511.08544, 2025.
28. Assran, M., et al., *V-jepa 2: Self-supervised video models enable understanding, prediction and planning*. arXiv preprint arXiv:2506.09985, 2025.
29. Dawid, A. and Y. LeCun, *Introduction to latent variable energy-based models: a path toward autonomous machine intelligence*. Journal of Statistical Mechanics: Theory and Experiment, 2024. **2024**(10): p. 104011.
30. Assran, M., et al. *Self-supervised learning from images with a joint-embedding predictive architecture*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
31. Chamroukhi, F., *Robust mixture of experts modeling using the t distribution*. Neural Networks, 2016. **79**: p. 20-36.
32. Zoph, B., et al., *St-moe: Designing stable and transferable sparse expert models*. arXiv preprint arXiv:2202.08906, 2022.
33. Jacobs, R.A., et al., *Adaptive Mixtures of Local Experts*. Neural Computation, 1991. **3**(1): p. 79-87.
34. Baldacchino, T., et al., *Variational Bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems*. Mechanical Systems and Signal Processing, 2016. **66-67**: p. 178-200.

35. Komatsuzaki, A., et al., *Sparse upcycling: Training mixture-of-experts from dense checkpoints*. arXiv preprint arXiv:2212.05055, 2022.
36. Cao, Z., et al., *Learning to rank: from pairwise approach to listwise approach*, in *Proceedings of the 24th international conference on Machine learning*. 2007, Association for Computing Machinery: Corvalis, Oregon, USA. p. 129–136.
37. Wang, X., et al., *CLIP-UP: A Simple and Efficient Mixture-of-Experts CLIP Training Recipe with Sparse Upcycling*. arXiv preprint arXiv:2502.00965, 2025.
38. Hayes, T., et al., *Simulating 500 million years of evolution with a language model*. *Science*, 2025. **387**(6736): p. 850–858.
39. Mysinger, M.M., et al., *Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking*. *Journal of Medicinal Chemistry*, 2012. **55**(14): p. 6582–6594.
40. Tran-Nguyen, V.-K., C. Jacquemard, and D. Rognan, *LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening*. *Journal of Chemical Information and Modeling*, 2020. **60**(9): p. 4263–4273.
41. Tan, H., et al. *Multi-Granular Contrastive Alignment and Fusion for Fragment-Enhanced Virtual Screening*. 2025.
42. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426, 2018.
43. Healy, J. and L. McInnes, *Uniform manifold approximation and projection*. *Nature Reviews Methods Primers*, 2024. **4**(1): p. 82.
44. Irwin, J.J., et al., *ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery*. *Journal of Chemical Information and Modeling*, 2020. **60**(12): p. 6065–6073.
45. Shivanyuk, A.N., et al., *Enamine real database: Making chemical diversity real*. *Chemistry today*, 2007. **25**(6): p. 58–59.
46. Landrum, G., et al., *rdkit/rdkit: 2025_03_1 (Q1 2025) Release*. Zenodo, 2025.
47. Wang, S., et al., *Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences*. *Journal of Chemical Information and Modeling*, 2020. **60**(4): p. 2044–2058.
48. Burley, S.K., et al., *Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive*, in *Protein Crystallography: Methods and Protocols*, A. Wlodawer, Z. Dauter, and M. Jaskolski, Editors. 2017, Springer New York: New York, NY. p. 627–641.
49. Varadi, M., et al., *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models*. *Nucleic Acids Research*, 2022. **50**(D1): p. D439–D444.
50. The UniProt, C., *UniProt: a hub for protein information*. *Nucleic Acids Research*, 2015. **43**(D1): p. D204–D212.
51. Morand, E., et al., *TYK2: an emerging therapeutic target in rheumatic disease*. *Nature Reviews Rheumatology*, 2024. **20**(4): p. 232–240.
52. Lupardus, P.J., et al., *Structure of the pseudokinase–kinase domains from protein kinase TYK2 reveals a mechanism for Janus kinase (JAK) autoinhibition*. *Proceedings of the National Academy of Sciences*, 2014. **111**(22): p. 8025–8030.
53. Wroblewski, S.T., et al., *Highly Selective Inhibition of Tyrosine Kinase 2 (TYK2) for the Treatment of Autoimmune Diseases: Discovery of the Allosteric Inhibitor BMS-986165*.

- Journal of Medicinal Chemistry, 2019. **62**(20): p. 8973-8995.
54. Chimalakonda, A., et al., *Selectivity Profile of the Tyrosine Kinase 2 Inhibitor Deucravacitinib Compared with Janus Kinase 1/2/3 Inhibitors*. Dermatology and Therapy, 2021. **11**(5): p. 1763-1776.
 55. Liu, J., R. Wu, and J.-X. Li, *TAAR1 as an emerging target for the treatment of psychiatric disorders*. Pharmacology & Therapeutics, 2024. **253**: p. 108580.
 56. Halff, E.F., et al., *Trace amine-associated receptor 1 (TAAR1) agonism as a new treatment strategy for schizophrenia and related disorders*. Trends in Neurosciences, 2023. **46**(1): p. 60-74.
 57. Zilberg, G., et al., *Molecular basis of human trace amine-associated receptor 1 activation*. Nature Communications, 2024. **15**(1): p. 108.
 58. Hu, Y.-X., et al., *Pseudouridine synthase 1 promotes hepatocellular carcinoma through mRNA pseudouridylation to enhance the translation of oncogenic mRNAs*. Hepatology, 2024. **80**(5).
 59. Huang, S., A.C. Wylder, and T. Pan, *Simultaneous nanopore profiling of mRNA m6A and pseudouridine reveals translation coordination*. Nature Biotechnology, 2024. **42**(12): p. 1831-1835.
 60. Carlile, T.M., et al., *mRNA structure determines modification by pseudouridine synthase 1*. Nature Chemical Biology, 2019. **15**(10): p. 966-974.
 61. Martinez, N.M., et al., *Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing*. Molecular Cell, 2022. **82**(3): p. 645-659.e9.
 62. Wang, F., et al., *Pseudouridine synthase 1-targeted therapy activates antiviral immunity to boost cancer immunotherapy*. Cell Reports, 2025. **44**(9).
 63. Chen, L., et al., *Pseudouridine synthase 1 promotes progression of hepatocellular carcinoma via mTOR and MYC signaling pathways*. Frontiers in Oncology, 2025. **15**: p. 1576651.
 64. Corrêa Veríssimo, G., R. Salgado Ferreira, and V. Gonçalves Maltarollo, *Ultra-Large Virtual Screening: Definition, Recent Advances, and Challenges in Drug Design*. Molecular Informatics, 2025. **44**(1): p. e202400305.
 65. Ruddigkeit, L., et al., *Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17*. Journal of Chemical Information and Modeling, 2012. **52**(11): p. 2864-2875.
 66. Buehler, Y., S. Javor, and J.-L. Reymond, *Sampling a GDB-20 Database of 32 Trillion Drug-Like Molecules by Generative Artificial Intelligence*. 2026.
 67. Singh, R., et al., *Contrastive learning in protein language space predicts interactions between drugs and protein targets*. Proceedings of the National Academy of Sciences, 2023. **120**(24): p. e2220778120.
 68. Lu, Y., et al., *\$ f \$-MICL: Understanding and Generalizing InfoNCE-based Contrastive Learning*. 2024.
 69. Rusak, E., et al., *Infonce: Identifying the gap between theory and practice*. 2024.
 70. Lei, H., et al., *M3-JEPA: Multimodal Alignment via Multi-gate MoE based on the Joint-Embedding Predictive Architecture*. 2024.
 71. Wang, Y., et al., *PubChem: a public information system for analyzing bioactivities of small*

- molecules*. Nucleic Acids Research, 2009. **37**(suppl_2): p. W623-W633.
72. Truchon, J.-F. and C.I. Bayly, *Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem*. Journal of Chemical Information and Modeling, 2007. **47**(2): p. 488-508.
73. Halgren, T.A., et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening*. Journal of Medicinal Chemistry, 2004. **47**(7): p. 1750-1759.
74. Trott, O. and A.J. Olson, *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.
75. Durrant, J.D. and J.A. McCammon, *NNScore 2.0: A Neural-Network Receptor-Ligand Scoring Function*. Journal of Chemical Information and Modeling, 2011. **51**(11): p. 2897-2903.
76. Ballester, P.J. and J.B.O. Mitchell, *A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking*. Bioinformatics, 2010. **26**(9): p. 1169-1175.
77. Stepniewska-Dziubinska, M.M., P. Zielenkiewicz, and P. Siedlecki, *Pafnucy-a deep neural network for structure-based drug discovery*. Bioinformatics, 2017. **1050**: p. 19.
78. Gao, B., et al., *Drugclip: Contrastive protein-molecule representation learning for virtual screening*. Advances in Neural Information Processing Systems, 2023. **36**: p. 44595-44614.
79. Zhou, G., et al. *Uni-mol: A universal 3d molecular representation learning framework*. in *The eleventh international conference on learning representations*. 2023.
80. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization*. 2014.
81. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery*. Nucleic Acids Research, 2012. **40**(D1): p. D1100-D1107.
82. Liu, T., et al., *BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data*. Nucleic Acids Research, 2025. **53**(D1): p. D1633-D1644.
83. Liu, T., et al., *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities*. Nucleic Acids Research, 2007. **35**(suppl_1): p. D198-D201.
84. Zhang, C., et al., *BioLiP2: an updated structure database for biologically relevant ligand-protein interactions*. Nucleic Acids Research, 2024. **52**(D1): p. D404-D412.
85. Wang, R., et al., *The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures*. Journal of Medicinal Chemistry, 2004. **47**(12): p. 2977-2980.
86. Liu, Z., et al., *Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions*. Accounts of Chemical Research, 2017. **50**(2): p. 302-309.
87. Grünberg, S., et al., *The structural basis of mRNA recognition and binding by yeast pseudouridine synthase PUS1*. PLOS ONE, 2023. **18**(11): p. e0291267.

Data availability

The data of ZEUS-3D is derived from ZINC20 (<https://zinc20.docking.org/>), ZINC22 (<https://cartblanche22.docking.org/>) and Enamine Real database (<https://enamine.net/compound-collections/real-compounds/real-database>).

Code availability

Computer codes of DrugJEPA are available at: <https://github.com/Saoge123/DrugJEPA>. Source codes of ULVS are available at: <https://github.com/Saoge123/Pipeline-of-ULVS-against-ZEUS-3D>

Acknowledgements

This work was supported by National Key R&D Program of China (grant no. 2023YFF1204905, S.Y.); the National Natural Science Foundation of China (grant nos. T2221004, S.Y.; 81930125, S.Y.; and 82273787, L.L.; 82404514, Y.J.; 82404414, C.H.; 82404409, W.S.; 22407098, L.Z.); the New Cornerstone Science Foundation; Major Project of Guangzhou National Laboratory (grant no. GZNL2024A01005); 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (grant nos. ZYXY21001, S.Y.; ZYGD23006, S.Y.); the Fellowship from the China Postdoctoral Science Foundation (2024M762258, L.Z.) and Sichuan University Postdoctoral Interdisciplinary Innovation Fund; We thank the Frontiers Medical Center, Tianfu Jincheng Laboratory for providing high-performance computing resources; We also thank the staff from beamlines BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility of the National Facility for Protein Science (Shanghai, China) for great support.

Contributions

S.Y. conceived and supervised the research and designed the experiments. S.Y. and Y.J. established and validated the model. Y.J. and M.D. performed data procession. C.H. and Q.H. carried out chemical synthesis. W.S, L.Z. and R.Y. performed the bioactivity assays. S.Y., Y.J., C.H., W.S., L.Z. and R.Y. analysed the data. S.Y. and Y.J. wrote the manuscript.

Supplementary Files

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

- [Sl.docx](#)