

Combinatorial docking and molecular generation to navigate over 100-billion molecules for prospective ligand discovery

Jianxiang Zhang^{1,2,†}, Chao Yang^{1,†}, Yang Zhang^{3,4,5,†}, Xiao Chen^{6,†}, Brandon Lam^{1,†}, Claire Bryant^{7,†}, Shabareesh Pidathala⁶, Yangzhi Wang^{1,2,8}, Yurii S. Moroz^{9,10,11}, Dmytro S. Radchenko⁹, Assaf Alon^{7,*}, Chia-Hsueh Lee^{6,*}, Zhe Zhang^{3,4,5,*}, Jiankun Lyu^{1,*}

¹The Evnin Family Laboratory of Computational Molecular Discovery, The Rockefeller University, New York, NY, USA.

²The David Rockefeller Graduate Program in Bioscience, The Rockefeller University, New York NY, USA.

³State Key Laboratory of Natural and Biomimetic Drugs, Department of Molecular and Cellular Pharmacology, School of Pharmaceutical Sciences, Peking University, Beijing, China

⁴Peking University-Yunnan Baiyao International Medical Research Center, Beijing, China

⁵School of Life Sciences, Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.

⁶Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.

⁷School of Medicine, Yale University, New Haven, CT, USA.

⁸Tri-Institutional PhD Program in Computational Biology and Medicine, New York, NY, USA.

⁹Enamine Ltd (www.enamine.net), Winston Churchill Street 78, Kyiv 02094, Ukraine.

¹⁰Taras Shevchenko National University of Kyiv, Volodymyrska Street 60, 01601, Kyiv, Ukraine.

¹¹Chemspace LLC (www.chem-space.com), Winston Churchill Street 85, Kyiv 02094, Ukraine.

† These authors contributed equally. * Corresponding authors

Abstract

Commercially available make-on-demand libraries now exceed 100 billion compounds, requiring over 50 years to screen on 2,000 CPU cores using conventional docking. We present two complementary approaches to address this challenge. CombiDOCK, a combinatorial docking framework, enables exhaustive screening at the 100-billion scale within 40 days. MINT-Dock, a generative framework, accelerates navigation of this space by integrating CombiDOCK with Monte Carlo Tree Search. Benchmarked on 46 diverse targets, CombiDOCK matched full-molecule docking accuracy, and MINT-Dock achieved a 4,800-fold enrichment over random selection. Compared with prior billion-scale brute-force campaigns against σ_2 , VMAT2, and VACHT, prospective CombiDOCK screens of the 100-billion-molecule library yielded higher hit rates and more potent ligands, while MINT-Dock achieved comparable outcomes across single- and multi-target objectives with >20-fold computational cost reductions. Docking-predicted poses of the best VACHT-binding compounds were confirmed by cryo-EM structures. These methods provide exhaustive and generative paths for navigating the trillion-molecule frontier of drug discovery.

Introduction

Recently, tangible virtual libraries fueled by diverse building block catalogs and robust parallel synthesis reactions (>80% success rates)^{1,2} have become major drivers in the expansion of commercially available chemical space. These fast-growing make-on-demand chemical libraries now exceed 100 billion structurally diverse and stereogenic molecules, offering an unprecedented opportunity for ligand and drug discovery. Recent prospective studies further suggest that well-predicted protein models can extend ligand discovery beyond experimentally determined structures, thereby broadening the range of targets accessible to virtual screening^{3–7}. Such large tangible libraries can only be screened computationally, and even then, require efficient and accurate methods^{8,9}. Many groups have leveraged these libraries for ligand discovery, including more than 45 prospective virtual screening campaigns,^{1,4,10–43} the majority of which rely on brute-force screening methods. When directly compared, larger libraries consistently identify more potent hits than screening smaller libraries,^{1,4,25,26,35,39,40} emphasizing the importance of utilizing larger libraries for virtual campaigns as they have become available.

Even with these advances, structure-based virtual screening (SBVS) remains fundamentally constrained by its underlying computational architecture. SBVS workflows typically rely on either on-the-fly ligand construction or pre-built conformer libraries.⁴⁴ Both workflows have distinct computational and storage demands; regardless, a 100-billion compound screen would take over 50 years on a 2000-core cluster for either approach (**Fig. 1A**, see details in **Table S1**),^{45–51} emphasizing that these strategies cannot feasibly scale to libraries on the order of 100 billion molecules. This scaling challenge is further amplified when discovery goals go beyond single-target binding to multi-target optimization, such as achieving selectivity over antitargets or polypharmacology across related receptors, goals that are central to many drugs and chemical probes^{52–61} but require separate exhaustive screens against each target, multiplying already prohibitive computational cost.

Despite the clear advantages of the larger libraries, their vastness poses substantial challenges for conventional docking approaches which struggle to scale beyond a few billion molecules. Several strategies have been developed to reduce this burden, including fragment-enrichment methods, such as Chemical Space Docking or V-SYNTHES,^{10,12} and machine-learning (ML) surrogate approaches.^{41,43,62,63} These approaches applied prospectively have identified sub-micromolar to low micromolar binders across multiple targets,^{10,12–14,16–20,42,43} all while screening far smaller libraries. Despite their efficiency, these approaches inevitably sacrifice accuracy: because they do not evaluate every molecule explicitly with the scoring function, relying instead on predicted or fragment-derived docking scores to filter the library, inaccuracies can propagate and many true positives are likely to be missed. In parallel, generative methods have emerged as an alternative way to explore only high-value regions of chemical space, enabling increasingly efficient discovery of novel and potent bioactives.^{64–83} However, current structure-based generative methods face a fundamental tradeoff among synthesizability,^{69,70,72} 3D pose prediction^{74,77,82} and generation throughput.^{71,81} As a result, relatively few studies have carried designs through synthesis and experimental testing at scale.

These challenges motivate us to develop two complementary computational goals for exploring over 100 billion-molecule make-on-demand chemical space (**Fig. 1B**). One is top-down, comprehensive virtual screening: to dock every molecule in a combinatorial library as accurately as possible without explicit full-molecule enumeration, thereby preserving the exhaustive logic of conventional docking while overcoming its scaling bottlenecks. The other is bottom-up, efficient navigation: to start from the underlying building blocks and reactions and use structure-guided

search to identify a much smaller set of high-value candidates while retaining explicit 3D evaluation of full molecules.

Here, we introduce two methods that achieve these complementary goals. For the top-down approach, we present CombiDOCK, a combinatorial docking approach that enables efficient, exhaustive screening of ultra-large chemical space by decomposing molecules into synthons and reconstructing full molecules on-the-fly during docking. Here, a synthon is defined as a building block fragment expressed in its product form, with reactive attachment points that specify how it can combine with other synthons under well-defined reaction rules.¹⁰ CombiDOCK is not a “fragment-enrichment then enumerate” workflow. It eliminates explicit full-molecule enumeration throughout both library representation and docking, enabling full-molecule-equivalent docking at >100 billion scale. For the bottom-up approach, building on CombiDOCK's combinatorial docking engine, we introduce MINT-Dock (MCTS Integrated Navigation Tool for Docking), a structure-based generative docking framework that couples Monte Carlo Tree Search (MCTS) with CombiDOCK to navigate make-on-demand chemical space by assembling reaction-compatible synthons and evaluating full molecules in 3D on the fly. While CombiDOCK exhaustively evaluates all possible combinations, MINT-Dock provides an adaptive, on-demand exploration that generates only the most promising products under explicit design objectives, including multi-target optimization for selectivity and polypharmacology, tasks that are otherwise computationally prohibitive at the ultra-large library scale.

To demonstrate the broad utility of CombiDOCK and MINT-Dock, we performed both retrospective testing against 46 diverse drug targets as well as prospective testing of up to 103-billion molecules in single and multi-target objective campaigns against the Sigma-2 (σ_2) receptor, vesicular monoamine transporter 2 (VMAT2) and vesicular acetylcholine transporter (VACHT). These three targets provide clinically relevant and methodologically stringent test cases of the method, and to our knowledge, these are the first docking campaigns against any target spanning over 100-billion compounds. To prospectively assess the hypothesis of whether bigger chemical libraries improve outcomes of ligand discovery, we synthesized and tested hundreds of top-ranking molecules from each screen. Hit rates and ligand affinities were compared with those from smaller library screens.

CombiDOCK efficiently addresses combinatorial virtual libraries through its modular design

Inspired by molecular fragmentation strategies,^{10,12} we designed CombiDOCK (implemented in DOCK6⁴⁵), a new docking method that builds ligands in the protein binding site by combining synthons using well-defined chemical transformations. Next, to further accelerate the navigation of ultra-large combinatorial chemical space, we introduce MINT-Dock, a structure-based generative framework that couples Monte Carlo Tree Search with CombiDOCK. Both methods operate on the same synthon-based representation of make-on-demand libraries and use the same CombiDOCK engine for evaluating full molecules in 3D, but differ in search strategy, with CombiDOCK enabling exhaustive exploration and MINT-Dock enabling adaptive navigation.

Using synthons and reactions to represent modern combinatorial libraries provides a far more compact and scalable representation than explicit enumeration of every possible product (**Fig. 1C**). This abstraction enabled us to encode 103 billion stereochemically rich and diverse molecules from Enamine REAL Space using a relatively small set of 2.3 million synthons and 304 reaction schemes. Because each building block can participate in one or more reaction schemes, we denote the combination of a building block–reaction pair as a synthon, yielding 6.3 million synthons in total. The resulting chemical space is partitioned into a two-component space (2.0

million synthons, 276 reactions; 5.5 billion molecules) and a three-component space (4.3 million synthons, 28 reactions; 97.7 billion molecules). For three-component reactions, because they involve a dependency in which one synthon serves as a linker connecting the other two, we pre-react the two smallest dependent synthon sets by merging the linker with one of the other synthons, generating a pseudo two-component representation compatible with both CombiDOCK and MINT-Dock, which currently combines only two synthons to assemble a full molecule. To enable structure-based modeling, synthons are capped according to the Markush structure of the reaction they participate in,¹⁰ preserving the local chemical environment of the final product while providing an alignment handle for assembling two 3D synthons into a full molecule (**Fig. 1C**). These capped synthons are then stereochemically expanded into 3D conformers, resulting in a diverse library of 103 billion molecules. This synthon-based representation, which more closely reflects the chemical environment of the final products than a building-block representation, plays an important role in both methods (**Fig. 1D**).

In comparison to the pre-sampling methods such as DOCK3⁵⁰ and its DOCK6 re-implementation,⁸⁴ CombiDOCK utilizes a computationally efficient on-the-fly approach capable of exhaustively evaluating the 50-fold larger combinatorial virtual libraries. (**Fig. 1E**). To avoid the redundancy of the pre-sampling methods, CombiDOCK decouples sampling: rather than pre-sampling the conformers of the entire compound at once, synthon 1 (Syn1) and synthon 2 (Syn2), which together constitute the full molecule, are sampled independently to generate conformational ensembles (**Fig. S1**). From here, the algorithm performs similar hierarchical growth to DOCK3, however, CombiDOCK decomposes the process into synthon-level growth steps: capped Syn1 is docked first, followed by capped Syn2, which is assembled under the corresponding reaction geometry. Because this procedure simply breaks down the same hierarchical growth process into synthon-defined stages, CombiDOCK is designed to produce results equivalent to full-molecule docking, while avoiding redundant sampling of identical substructures across the library. The power of CombiDOCK lies in its ability to scale with the growing size of combinatorial libraries. Instead of enumerating and docking every possible full molecule, which becomes intractable at $>10^{10}$ scale, CombiDOCK docks each synthon fragment once to generate its binding poses within the receptor. These cached synthon poses are then reused to assemble full molecules on-the-fly, ensuring that fragment-level and product-level docking remain consistent. As illustrated in the **Fig. 1E**, Syn1 is first docked into the binding site to produce an ensemble of poses. Partner synthons (Syn2s) are then introduced and combined with the cached Syn1 poses to yield full Syn1-Syn2 products, which are subsequently ranked by docking score. The process is repeated with Syn2 as the first fragment, ensuring both orientations are explored, after which all possible Syn1-Syn2 combinations can be reconstructed without redundant calculations.

MINT-Dock: Structure-based generative docking with Monte Carlo Tree Search

While CombiDOCK exhaustively evaluates the full combinatorial space, MINT-Dock uses a Monte Carlo Tree Search (MCTS) architecture to traverse this space while remaining strictly within synthesizable, reaction-defined combinations. Synthon docking scores are precomputed once using DOCK6 HDB docking on reaction-capped synthons⁸⁵ (see **Methods**), and these scores guide early-stage exploration without requiring product-level docking at every step. The search is organized as a three-layer tree. In the first two layers, each node corresponds to a synthon, and each path specifies a unique combination of reaction-compatible synthons that can be merged into a product node in the third (final) layer. As illustrated in **Fig. 1F**, the tree expands in parallel each cycle, typically branching to 500 to 1,000 synthon₁ choices at the first layer and 20 to 50 synthon₂ choices at the second layer, prioritized by synthon scores. When a branch reaches the product layer, the selected synthons are reacted *in silico* to form full molecules, and 10,000 to 50,000 products are docked in parallel using CombiDOCK to obtain full-molecule scores (**Fig. 1F**).

For each synthon combination, only the top-ranked stereoisomer is retained to promote diversity in the generated pool. The resulting product-level score is used as the MCTS reward and is backpropagated to the corresponding synthon nodes, updating their expected values and biasing subsequent rollouts toward higher-reward branches. Repeating this process over 200 to 400 cycles separately on two-component and three-component space typically yields over 8 million docked, synthesizable molecules per campaign, which are then passed through downstream triage (score, ligand strain, novelty to knowns, and interaction filters), followed by visual inspection and compound synthesis (**Fig. 1F**). For multi-target tasks such as selectivity or polypharmacology, MINT-Dock defines a design function that maps per-target docking scores to a single scalar reward used for node selection and backpropagation. For selectivity, the design function increases with the score of the intended on-target and decreases with the score of the off-target, whereas for polypharmacology, it increases with the scores of both targets. During search, the design function is evaluated using the precomputed synthon docking scores for each target, and during backpropagation it is evaluated using the full-molecule docking scores for each target.

Next, using σ_2 and VMAT2 as test cases, we evaluated two features that improve MINT-Dock search performance. We first compared the accumulation of well-scored molecules with and without backpropagation across rollouts and found that backpropagation consistently increased the yield of well-scoring molecules across multiple score cutoffs, and the advantage became more pronounced as the number of cycles increased (**Fig. S2** and **S3**, see details in **Methods**). We then compared MINT-Dock to an ML-guided MCTS baseline (“ML-MCTS”, adapted from SynthemoI⁶⁷) to assess whether coupling MCTS to physics-based docking improves generation relative to an ML surrogate. In both targets, MINT-Dock produced substantially more well-scored molecules above any given score threshold, yielding higher enrichment across thresholds (**Fig. S4**, see details in **Methods**).

Benchmarking CombiDOCK/MINT-Dock against full-molecule docking on DUDE-Z benchmark set

To assess the generalizability of CombiDOCK’s performance, we evaluated it against 43 diverse protein targets from the widely used DUDE-Z benchmark set (dudez.docking.org),⁸⁶ along with 3 prospectively tested targets, for a total of 46. These targets span diverse protein classes (enzymes, proteases, GPCRs, kinases, nuclear receptors, among others) and were curated on the basis of the availability of complete resolved structures and known ligands, as well as a broad range of preferred ligand charge states. This benchmark set ensures that performance is measured across a broad range of pharmacologically relevant targets with varied binding site characteristics (see **Table S2**) and enables direct comparison across different docking engines^{87–90} (see **Table S3**). Each target includes experimentally tested actives and computationally generated decoys that are property-matched to the actives but topologically dissimilar. To represent this benchmarking set in synthon space for direct comparison to conventional full-molecule docking, each full molecule was systematically decomposed into its constituent synthons and capped (**Fig. 2A**). The resulting capped synthon libraries were then organized by reaction type, serving as reusable inputs for CombiDOCK docking. Across the 46 targets, docking scores between full-molecule docking and CombiDOCK were highly correlated, with a mean Pearson correlation coefficient (R) of 0.82 (**Fig. 2B**). Enrichment of known actives over decoys was nearly identical, with a mean $\Delta\log$ -adjusted area under the curve ($\Delta\log\text{AUC}$) of 0.97 between full-molecule docking and CombiDOCK (**Fig. 2C**). For the three prospective targets: σ_2 receptor, VMAT2, and VACHT, docking scores were again strongly correlated ($R = 0.91$ for σ_2 , 0.82 for VMAT2, and 0.91 for VACHT), with nearly identical enrichment values (**Fig. 2B-C**). These findings

demonstrate that CombiDOCK achieves accuracy comparable to full-molecule docking while offering a fundamentally more scalable framework.

Next, to quantify the computational efficiency of MINT-Dock, we benchmarked it on the same DUDE-Z set for its ability to enrich well-scored molecules relative to docking a random subset from the same library (see details in **Methods**). For each of the 46 targets, we computed an adjusted enrichment that compares the number of molecules exceeding a given score threshold between MINT-Dock and the random subset while accounting for differences in sample size and upweighting more favorable scores (see details in **Methods**). **Fig. 2D** summarizes the adjusted enrichment at the designated score threshold across all 46 targets, spanning orders of magnitude and resulting in a mean improvement of 950-fold over random selection when the random subset was docked with CombiDOCK. We report this comparison because CombiDOCK is optimized for combinatorial libraries and is the same product-level docking engine used inside MINT-Dock, making it a particularly stringent and method-matched comparison. In parallel, we also benchmarked against one of the fastest conventional full-molecule docking methods, DOCK6 HDB, to provide a more general comparison for library-agnostic reference. Under this baseline, the mean enrichment increased by 4,810-fold (**Fig. S5A**). A more fine-grained enrichment analysis on three representative targets showed that MINT-Dock consistently produced more molecules above any given score cutoff than random sampling, with the strongest enrichment at the most stringent thresholds, indicating that the search preferentially concentrates docking on productive regions of chemical space (**Fig. S5B**).

Docking scores continue to improve with library size at the >100-billion-molecule scale

A central question for make-on-demand combinatorial libraries is whether expanding their size continues to yield more favorable molecules or whether performance eventually saturates beyond 5 billion molecules. This question has been difficult to address because the computational cost of exhaustive screening at this scale is prohibitive for conventional methods. CombiDOCK overcomes this barrier, enabling brute-force screening of entire libraries and thus allowing this scaling question to be tested directly. As a proxy, we examined how docking scores improve with increasing library size.⁹¹ Although docking scores are imperfect predictors of binding affinity, they have been shown to correlate with experimental hit rates,²⁵ and they remain the primary criterion for compound selection in virtual screening.

Before scaling to 100+-billion-molecule libraries, we first optimized the two sampling parameters that govern CombiDOCK's speed-accuracy tradeoff (**Fig. S6**, see details in **Methods**). With these parameters fixed, we docked progressively larger subsets of Enamine REAL Space against three prospective targets: σ_2 receptor, VMAT2, and VACHT to test scaling behavior. From the largest enumerated sets (56 billion for σ_2 , 103 billion for VMAT2, and 103 billion for VACHT), we sampled 20 random subsets spanning approximately 10^7 to 10^{11} molecules in half-logarithmic increments. For each subset, we analyzed the top-ranked molecules: 300,000 for σ_2 and 1 million for VMAT2 and VACHT, calculating their minimum, maximum, median, and quartile docking scores. As shown in **Fig. 2E**, the docking scores of top-ranking molecules improved monotonically and approximately log-linearly with increasing library size across all three targets. This improvement was consistent across quartiles, indicating that the entire distribution of high-ranking compounds steadily shifts toward better fits as the library grows. Importantly, no signs of score saturation were observed, suggesting that even at scales beyond 10^{11} molecules, further expansion continues to yield better-fitting candidates. The computational cost of these screens was substantial but tractable with modern resources: using a cluster of 2,000 CPU cores, docking required 40 cluster days for σ_2 (56B molecules), 33 cluster days for VMAT2 (103B molecules), and 26 cluster days for VACHT (103B molecules). These results highlight that both the quality of top-ranking

molecules and the feasibility of screening at trillion scale improve hand-in-hand, supporting the case for continued library expansion.

Prospective testing on combinatorial docking and molecular generation

To test whether these computational gains translate into ligand discovery with improved outcomes, we conducted prospective screens using conventional docking method DOCK3, CombiDOCK and MINT-Dock across both well-studied and challenging targets for structure-based drug design. Here, we sought to evaluate whether guided search of the 100 billion molecule library continues to yield novel and more potent chemotypes with substantially less computational cost relative to conventional docking methods. We address these questions target by target, beginning with the σ_2 receptor.

Prospective combinatorial docking and molecular generation against the σ_2 receptor

To address this question in a setting with a well-established large-scale docking baseline, we chose σ_2 as benchmark target. Because σ_2 has a previous prospective campaign that screened 490 million molecules,²⁷ this enables an apples-to-apples comparison across both library sizes and search strategies. In this earlier effort, 138 high-scoring compounds from top 300,000 were tested in a competition binding assay with [³H]-DTG and the known σ_2 ligand PB28 was included as a positive control. The hit rate was 51% (70/138 displaced >50% [³H]-DTG at 1 μ M) (**Fig. 3A-B**). The top 21 potent ligands exhibiting K_i values of 30 nM to the low μ M range (6 with K_i < 50 nM, none < 10 nM) (**Fig. 3C-D**). This campaign serves as the baseline for evaluating both CombiDOCK and MINT-Dock below.

CombiDOCK exhaustive screen (56 billion molecules). Using CombiDOCK, we screened 56 billion compounds from the REAL Space virtual library against the same σ_2 receptor crystal structure in complex with cholesterol (PDB ID: 7MFI).²⁷ On a 2,000-CPU-core cluster, docking required 40 cluster-days, despite the 100-fold increase in library size relative to our previous conventional docking campaign, in which 490 million make-on-demand compounds from the ZINC20 library² were screened using DOCK3²⁷ in three cluster-days. To match the selection depth used in the prior 490-million-compound campaign, 122 high-ranking molecules were synthesized and tested from the top 300,000 scoring molecules. For consistency, we used the same post-docking filters for selecting molecules from the previous 490 million molecule campaign²⁷ (detailed in the **Methods**) and the same binding assay conditions, including PB28 as a positive control. This yielded a hit rate of 67% (82/122 displaced >50% [³H]-DTG). Strikingly, the top 18 actives spanned affinities ranging from 6 to 473 nM, with 12 ligands < 50 nM and three < 10 nM (**Fig. 3C-D**). The highest affinity molecule CSMS01699433672 has a K_i value of 6 nM, whose docked pose forms key salt bridge interaction with D29 as well as an additional hydrogen bond with T110 (**Fig. 3E**). Its dose-response curve is presented in **Fig. 3F**.

MINT-Dock guided search (103 billion molecules). Starting from the synthons of the 103 billion Enamine REAL Space, MINT-Dock generated 10.3 million compounds for downstream processing in 4 days with approximately 100 CPU cores. From the top 300,000 ranked compounds from MINT-Dock screen, 206 compounds were synthesized and tested. Eighty-three compounds displaced more than 50% of [³H]-DTG (**Fig. 3A**), corresponding to a hit rate of 40% (**Fig. 3B**). Among the top 18 compounds ranked by [³H]-DTG displacement, 7 of 18 hits achieved K_i below 50 nM, with the best compound possessing a K_i of 14 nM (**Fig. 3C-D**), comparable to PB28 (K_i = 33 nM measured in the same assay), whereas in the prior campaign 6 of 21 compounds achieved K_i < 50 nM and the best K_i was 30 nM (**Fig. 3C-D**). The most potent

compound's docked pose forms key salt bridge interaction with D29 as well as an additional hydrogen bond with Q77 (**Fig. 3E**). Its dose-response curve is presented in **Fig. 3F**.

Comparative analysis across σ_2 campaigns. Across the three σ_2 campaigns, screening larger chemical space improved both hit rate and potency distribution. CombiDOCK achieved the highest hit rate, 67% versus 51% in the prior 490-million baseline, and improved the best potency from 30 nM to 6 nM, making its top ligand also about 5-fold more potent than the well-characterized σ_2 ligand PB28 ($K_i = 33$ nM). MINT-Dock achieved a 40% hit rate with a best K_i of 14 nM, thus delivering comparable or improved potency relative to the prior 490-million campaign while requiring 22-fold less docking time and approximately 240-fold less storage (1 TB for synthon conformers versus 240 TB for full-molecule conformers).

Prospective combinatorial docking and molecular generation against SLC transporters VMAT2 & VACHT

We next asked whether the scaling effect observed with σ_2 generalized to challenging neurotransmitter solute carrier (SLC) transporters.^{92–97} We selected VMAT2 and VACHT as a paired test case because they are in the same family, share similar methodological constraints, and are both highly clinically relevant. They also adopt the major facilitator superfamily (MFS) fold, a widely conserved structural class within the SLC superfamily that remains underexplored by large-scale virtual docking. To our knowledge, few prospective docking campaigns at the >1-billion-compound scale, if any, have been reported for MFS transporters. Both are challenging targets for structure-based ligand discovery due to their extensive conformational dynamics,^{98–101} and historically limited structural information available for SLC proteins. Only recently have their high-resolution experimental structures been determined,^{101,102} enabling structure-based discovery of new chemotypes. This limited structural knowledge in the protein family makes pose prediction especially challenging. Notably, although AlphaFold3 (AF3) performs comparably to DOCK3 for σ_2 (RMSD < 2 Å), its performance does not generalize to VMAT2 or VACHT.¹⁰³ Both DOCK3 and CombiDOCK accurately reproduced the cryo-EM binding poses of known ligands (RMSD < 2 Å), whereas AF3 failed to generate qualitatively correct binding modes (**Fig. S7A-B**). This trend held across a broader panel of 154 SLC transporters, revealing a systematic limitation of AF3: restricted generalizability to protein-ligand complexes underrepresented or absent in its training data (**Fig. S7C**). In contrast, physics-based approaches such as DOCK3 and CombiDOCK maintained robust performance across these out-of-distribution targets, underscoring why advancing scalable physics-based docking remains conceptually important even for ligandable targets.

Beyond their methodological challenges, both targets present pressing therapeutic needs that motivate the discovery of new chemotypes beyond their existing privileged scaffolds. The Vesicular Monoamine Transporter 2 (VMAT2, SLC18A2) is an antiporter that actively loads monoamine neurotransmitters from the cytosol into synaptic vesicles.¹⁰⁴ It is a key drug target for movement disorders, including tardive dyskinesia and chorea associated with Huntington's disease.^{105–107} Although VMAT2 is a clinically validated target, important therapeutic gaps remain.¹⁰⁸ Inter-patient variability in CYP2D6 metabolism can alter exposure and tolerability of current VMAT2 inhibitor scaffolds,¹⁰⁹ motivating the search for new chemotypes with improved pharmacological properties. The Vesicular Acetylcholine Transporter (VACHT, SLC18A3) performs an analogous function, specifically loading acetylcholine into synaptic vesicles in cholinergic neurons,^{110,111}. While it is a promising diagnostic target for Alzheimer's disease, existing vesamicol-class ligands lack selectivity over σ receptors, limiting their clinical imaging utility.¹¹² For both targets, structure-based screening of ultra-large chemical space offers a path

to move beyond these long-established scaffold classes and identify fundamentally new chemotypes that can potentially address these pressing clinical and translational needs.

We conducted prospective screens against VMAT2 (PDB ID: 8T69, tetrabenazine-bound complex)¹¹³ and VACHT (PDB ID: 8ZMR, vesamicol-bound complex)¹⁰¹ across libraries of increasing size (3.6 million (DOCK3.8), 1.9 billion (DOCK3.8), and 103 billion (CombiDOCK and MINT-Dock)). We present the VMAT2 results below, followed by VACHT in the next section.

DOCK3 screens (3.6 million and 1.9 billion molecules). Using full-molecule docking with DOCK3, we screened 3.6 million and then 1.9 billion compounds against VMAT2. Full-molecule docking of the 1.9-billion-compound library required 11 cluster-days on a 2,000-CPU-core cluster. Across these screens, we selected top-scoring compounds for experimental testing from the top-ranked subset (top 100K for 3.6M; top 1M for 1.9B) using the same post-docking filters detailed in **Methods**. We used a top-100K cutoff for the 3.6-million screen, rather than top 1M, because selecting 1M compounds from a 3.6M library would be both impractical and unnecessary; instead, we capped selection at the top 3% of the ranked list, already a generous cutoff given that prior in-stock screens typically select compounds from the top 0.8-1.0%.^{35,39} These campaigns resulted in the synthesis and testing of 65 and 111 compounds, respectively. In a [³H]-dopamine ([³H]-DA) uptake assay at 10 μ M, VMAT2 hit rates were 23% in the 3.6-million screen (15/65 displaced >50% [³H]-DA) and 27% in the 1.9-billion screen (30/111 displaced >50% [³H]-DA) (**Fig. 3G-H**). In the 3.6-million screen, IC₅₀ values ranged from 2 μ M to 7 μ M (six with IC₅₀ > 1 μ M, none with IC₅₀ between 200 nM and 1 μ M, none with IC₅₀ < 200 nM). In the 1.9-billion screen, IC₅₀ values ranged from 230 nM to 2 μ M (one with IC₅₀ > 1 μ M, ten with IC₅₀ between 200 nM and 1 μ M, none with IC₅₀ < 200 nM) (**Fig. 3I-J**).

CombiDOCK exhaustive screen (103 billion molecules). CombiDOCK was used to screen 103 billion molecules against VMAT2, requiring 33 cluster-days on a 2,000-CPU-core cluster. This campaign resulted in the synthesis and testing of 109 compounds from the top 1 million ranked molecules. The hit rate further improved to 47% (51/109 displaced >50% [³H]-DA) (**Fig. 3G**). IC₅₀ values ranged from 125 nM to 1 μ M (one with IC₅₀ = 1 μ M, nine with IC₅₀ between 200 nM and 1 μ M, two with IC₅₀ < 200 nM). The top compound has IC₅₀ of 125 nM (**Fig. 3I-J**), whose docked pose preserves all key interactions with the VMAT2 receptor (E312, R189, F135) (**Fig. 3K**). Its dose-response curve is presented in **Fig. 3L**.

MINT-Dock guided search (103 billion molecules). Starting from the synthons of the 103 billion Enamine REAL Space, MINT-Dock generated 15.6 million docked, synthesizable compounds for downstream processing in 4 days using approximately 95 CPU cores. 108 compounds were selected and tested from the top 1 million ranked molecules. Thirty-seven compounds displaced more than 50% of [³H]-DA (**Fig. 3G**), corresponding to a 34% hit rate (**Fig. 3G-H**). We further characterized the top 13 hits with dose-response curves, obtaining an overall potency profile in which 5 of 13 compounds had IC₅₀ < 1 μ M and 12 of 13 had IC₅₀ < 2 μ M (**Fig. 3I-J**). The most potent compound had an IC₅₀ of 126 nM, whose docked pose similarly preserves all key interactions with the VMAT2 receptor (E312, R189, F135) (**Fig. 3K**).

Comparative analysis across VMAT2 campaigns. Across the four VMAT2 campaigns, screening larger chemical space improved both hit rate and potency distribution. CombiDOCK achieved the highest hit rate, 47% versus 23% and 27% in the 3.6M and 1.9B brute-force campaigns, and improved the best potency from 230 nM in the 1.9B screen to 125 nM, approaching the potency of the FDA-approved VMAT2 inhibitor tetrabenazine (TBZ; IC₅₀ = 92 nM, **Fig. 3L**). MINT-Dock achieved a 34% hit rate, exceeding the prior 1.9B brute-force campaign, with a best IC₅₀ of 126 nM, also slightly better than the best compound from the 1.9B screen and

likewise comparable to TBZ, while requiring 64-fold less docking computation and approximately 930-fold less storage.

Having established that both CombiDOCK and MINT-Dock improve ligand discovery outcomes for VMAT2, we next turned to VACHT, the second vesicular SLC transporter in our paired test case. As described above, existing vesamicol-class ligands, the privileged scaffold for VACHT, lack selectivity over σ receptors, limiting their clinical utility as imaging agents for Alzheimer's disease.^{114,115} VACHT therefore presents not only a single-target discovery challenge but also a multi-target optimization problem: identifying compounds that are potent against VACHT while selectively avoiding σ_2 activity, or alternatively, identifying dual-active compounds for polypharmacology applications. We conducted progressive single-target screens (3.6M, 1.9B, 103B) using DOCK3 and CombiDOCK, and additionally leveraged MINT-Dock's multi-target capabilities for explicit selectivity and polypharmacology campaigns against VACHT and σ_2 .

DOCK3 screens (3.6 million and 1.9 billion molecules). Using full-molecule docking with DOCK3, we screened 3.6 million and then 1.9 billion compounds against VACHT (PDB ID: 8ZMR, vesamicol-bound complex).¹⁰¹ These campaigns resulted in the synthesis and testing of 64 and 108 compounds, respectively. In a [³H]-vesamicol competition binding assay at 7.5 μ M, hit rates were 25% in the 3.6-million screen (16/64 displaced >50% [³H]-vesamicol) and 28% in the 1.9-billion screen (30/108 displaced >50% [³H]-vesamicol) (**Fig. 4A**). In the 3.6-million screen, IC₅₀ values ranged from 200 nM to the low μ M range (two with IC₅₀ > 1 μ M, three with IC₅₀ between 100 nM and 1 μ M, none with IC₅₀ < 100 nM). In the 1.9-billion screen, IC₅₀ values ranged from 103 nM to the low μ M range (two with IC₅₀ > 1 μ M, 13 with IC₅₀ between 100 nM and 1 μ M, none with IC₅₀ < 100 nM) (**Fig. 4B**).

CombiDOCK exhaustive screen (103 billion molecules). CombiDOCK was used to screen 103 billion molecules against VACHT, requiring 26 cluster-days on a 2,000-CPU-core cluster. This campaign resulted in the synthesis and testing of 114 compounds from the top 1 million ranked molecules. The hit rate further improved to 56% (64/114 displaced >50% [³H]-vesamicol) (**Fig. 4A**). IC₅₀ values ranged from 12 nM to 847 nM (none with IC₅₀ > 1 μ M, 12 with IC₅₀ between 100 nM and 1 μ M, three with IC₅₀ < 100 nM). Notably, sub-200 nM ligands emerged only at the 103-billion scale. The top compound has an IC₅₀ value of 12 nM (**Fig. 4B**).

MINT-Dock selectivity campaign (VACHT selective over σ_2). Beyond single-target screening, MINT-Dock's multi-target capabilities enabled explicit optimization for VACHT selectivity over σ_2 : a key clinical objective that cannot be addressed by exhaustive single-target screening alone. MINT-Dock generated 12.9 million compounds for downstream processing in 3.5 days with approximately 110 CPU cores. With the same post-docking filtering criteria used for the CombiDOCK and prior campaigns (detailed in **Methods**), 114 were synthesized and tested for VACHT activity. In a [³H]-vesamicol displacement assay, 49 of 114 compounds displaced >50% of radioligand (**Fig. 4C**), corresponding to a 43% VACHT hit rate. To assess selectivity, we then tested these VACHT hits against σ_2 using a [³H]-DTG displacement assay. Forty-five of 49 compounds displaced <50% of [³H]-DTG (**Fig. 4D**), indicating that 92% of VACHT actives were σ_2 selective under this criterion. Overall, 45 of 114 synthesized compounds satisfied the selectivity profile (VACHT active and σ_2 inactive), yielding an overall hit rate of 40%. We advanced 16 selectivity hits for dose-response curves against both targets. These compounds showed VACHT potencies spanning the sub-100 nM to low-micromolar range, with the two most potent compounds achieving VACHT IC₅₀ = 86 nM (Z9424360192) and 100 nM (Z2406174633) (**Fig. 4E**). In parallel σ_2 profiling, most compounds remained weak binders, with 9 of 16 exhibiting σ_2 Ki > 10 μ M, consistent with the intended selectivity task (**Fig. 4F**).

MINT-Dock polypharmacology campaign (VACHT plus σ_2). For the polypharmacology campaign, MINT-Dock generated 8.8 million compounds for downstream processing in 3 days with approximately 75 cores. 53 compounds were synthesized and tested. In the [3 H]-vesamicol displacement assay, 38 of 53 compounds displaced >50% of radioligand (**Fig. 4G**), corresponding to a 72% VACHT hit rate. We then evaluated these VACHT hits against σ_2 , and 21 of 38 displaced >50% of [3 H]-DTG (**Fig. 4H**), corresponding to a 55% σ_2 hit rate among VACHT actives and yielding an overall polypharmacology hit rate of 21 of 53 (40%). Across the polypharmacology-tested compounds, higher VACHT displacement was associated with higher σ_2 displacement (**Fig. S8A**), and vice versa (**Fig. S8B**), consistent with the expectation that the polypharmacology objective enriches dual-target binders rather than producing VACHT and σ_2 activity independently by chance. Among the top 16 polypharmacology hits followed up with dose-response measurements, IC₅₀ values on VACHT spanned from single-digit nanomolar to low micromolar, with the most potent compound reaching VACHT IC₅₀ = 10 nM (Z9506560893) and multiple additional compounds in the sub-200 nM range (**Fig. 4I**). Consistent with the dual-target task, σ_2 binding for this set also shifted into a more potent range, with Ki values extending to tens of nanomolar for the strongest binders (**Fig. 4J**).

Comparative analysis across VACHT campaigns. Across the five VACHT campaigns, screening larger chemical space improved both hit rate and potency distribution, while MINT-Dock further enabled objective-specific multi-target discovery (**Fig. 4K,L**). CombiDOCK achieved the highest single-target hit rate, 56% versus 25% and 28% in the 3.6M and 1.9B brute-force campaigns, respectively, and improved the best potency from 103 nM in the 1.9B screen to 12 nM, a roughly 150-fold improvement over [+-]vesamicol in the same assay. MINT-Dock achieved higher VACHT hit rates than the 1.9B brute-force campaign in both multi-target modes, 43% for the selectivity campaign and 72% for the polypharmacology campaign, with more than 80-fold less computational cost. Its best compounds reached 86 nM in the selectivity campaign and 10 nM in the polypharmacology campaign, while steering discovery toward distinct endpoints: the selectivity campaign yielded compounds with significantly higher selectivity indices, whereas the polypharmacology campaign produced dual VACHT/ σ_2 binders (**Fig. 4M,N**). Unexpectedly, even without explicit selectivity optimization, the top CombiDOCK compound, CSMS00228273820, exhibited a Ki of >10 μ M for σ_2 , corresponding to >800-fold selectivity over σ_2 . This exceeds the selectivity of [+-]vesamicol (**Fig. 5A**), prior reported values for ($\pm$)-benzovesamicol (73-fold) and spiroindolines (<8-fold),¹¹⁵ and the best selectivity hit from the MINT-Dock campaign (116-fold, **Fig. 5A**). Together, these hit-rate, potency, and selectivity profiles show that MINT-Dock can steer discovery toward qualitatively distinct multi-target endpoints, while CombiDOCK can independently uncover highly selective compounds through exhaustive coverage of chemical space.

Structural validation of CombiDOCK/MINT-Dock-predicted poses. Unlike conventional docking that operates on pre-enumerated full molecules, CombiDOCK assembles synthons into full molecules on the fly, completely eliminating the enumeration step. Because the docking poses are constructed through this synthon-staged assembly rather than sampled from precomputed full-molecule conformers, it is important to validate that the resulting predicted poses are structurally faithful. We therefore selected the top compound from each method: the most potent from the CombiDOCK exhaustive screen and the most selective from the MINT-Dock selectivity campaign, and we determined cryo-EM structures of VACHT bound to each molecule.

For the CombiDOCK-derived compound CSMS00228273820 (IC₅₀ = 12 nM, **Fig. 5A**), we obtained a 3.6 Å-resolution cryo-EM structure of its complex with VACHT in an outward-open state (**Fig. 5B** and **Fig. S9** and **Table S4**). The local resolution around the ligand-binding site reached approximately 3.3 Å, allowing unambiguous determination of the ligand pose and surrounding

side-chain conformations (**Fig. S10A**). The experimentally resolved pose closely matched the CombiDOCK prediction (RMSD = 1.4 Å; **Fig. 5C**), confirming bidentate interactions with D398. In contrast, AF3 failed to accurately predict the binding mode (best RMSD of 3.8 Å; **Fig. S11A**). Combined with >800-fold selectivity over σ_2 , these data highlight its promise for future therapeutic development (**Fig. 5A**, bottom panel).

For the MINT-Dock-derived selective compound Z9424360192 (IC_{50} = 86 nM), we determined a 3.7 Å-resolution cryo-EM structure of its VACHT complex (**Fig. S12** and **Table S4**). Local resolution near the ligand-binding site was similarly high (3.3 Å), enabling clear pose assignment (**Fig. S10A**). Superimposition with the MINT-Dock prediction showed close agreement (RMSD = 1.9 Å; **Fig. 5C**), verifying bidentate interactions with D398 and additional bidentate contacts with N302. In this case, the top-confidence AF3 pose was also close to the experimental structure, with an RMSD of 2.2 Å (**Fig. S11B**). Across the three experimental VACHT structures (vesamicol, CSMS00228273820, and Z9424360192), the 33% success rate of AF3 pose reproduction followed the same overall accuracy trend shown in **Fig. S7C**.

The independent cryo-EM validation of both predicted poses confirms the structural fidelity of the shared combinatorial docking framework. We further performed MicroScale Thermophoresis (MST) to validate the binding of CSMS00228273820 to VACHT and to confirm the functional significance of the two key interaction residues, D398 and N302 (**Fig. S10B**).

Analogue optimization. To assess the capability of the combinatorial strategy for analogue optimization, we pursued analogue searches around top hits from both campaigns. For VMAT2, we applied a synthon-based approach in which close analogues were rapidly identified by searching matching synthons around parent hits. Across four primary CombiDOCK hits, 103 analogues were synthesized and tested, of which 72 (70%) exhibited >50% inhibition in the same primary assay described above. Dose-response follow-up identified more potent analogues for three of the four ligands, with potency improvements of up to 9-fold and a best IC_{50} of 77 nM (**Fig. S13A-E**), comparable to the FDA-approved VMAT2 inhibitor tetrabenazine. For VACHT, we searched Enamine REAL Space with [Small World](#) for close analogs around the top CombiDOCK hit and two MINT-Dock selective hits. These analogue searches yielded high hit rates and identified affinity improvements while maintaining selectivity. Notably, two series converged on the same optimized scaffold, which reached a VACHT IC_{50} of 8 nM and 649-fold selectivity over σ_2 (**Fig. S13F**), exceeding prior reported selectivity values for ($\pm$)-benzovesamicol (73-fold) and spiroindolines (<8-fold).¹¹⁵ Across five optimization series spanning both methods, three showed 8- to 12-fold improvements, supporting that the workflow can efficiently progress from initial hit discovery to improved potency without bespoke synthesis campaigns.

Scaffold novelty analysis. Bemis-Murcko scaffold analysis¹¹⁶ across the σ_2 , VMAT2, and VACHT campaigns showed that screening previously intractable chemical space with either CombiDOCK or MINT-Dock consistently uncovered previously unseen chemotypes (**Fig. S14**). For σ_2 , scaffold overlap between the smaller DOCK3 screen and the larger CombiDOCK and MINT-Dock screens was minimal, with only at most two pairwise shared scaffolds and just one scaffold shared across all three campaigns (**Fig. S14A-B**). Similarly low overlap was observed for VMAT2 and VACHT between the smaller in-stock/ZINC library screens performed with DOCK3 and the larger CombiDOCK and MINT-Dock campaigns, where shared scaffolds were rare and most tested scaffolds were unique to each screen (**Fig. S14C-H**). Even when CombiDOCK and MINT-Dock searched the same underlying chemical space, they recovered largely distinct scaffold sets, indicating that exhaustive and guided strategies sample different productive regions of that space (**Fig. S14D,H**). Together, this structural diversity across methods and objectives

confirms that both exhaustive and guided approaches reveal novel chemotypes at previously inaccessible scale.

Fragment-enrichment based virtual screening versus CombiDOCK/MINT-Dock: a comparative analysis

Both CombiDOCK and MINT-Dock operate on synthon-based chemical space, a representation shared by fragment-enrichment methods such as V-SYNTHES. However, the three approaches differ fundamentally in how they exploit this representation. Fragment-enrichment methods dock individual synthons to identify high-scoring anchor fragments, enumerate only products containing those anchors, and then dock the reduced product set at the full-molecule level. CombiDOCK instead docks capped synthons once and recombines cached poses to assemble and score all full molecules on the fly. MINT-Dock further couples this combinatorial docking engine with MCTS-guided search to concentrate docking on the most promising regions of chemical space, and its multi-target optimization additionally enables selectivity-driven optimization: a task that fragment-enrichment methods cannot address. We evaluated how fragment-enrichment performs relative to both methods, first retrospectively in recall of top-scoring molecules and then prospectively in recovery of experimental actives reported in this study.

To evaluate the fidelity of our approach with combinatorial libraries, we retrospectively benchmarked CombiDOCK against full-molecule docking across σ_2 , VMAT2, and VACHT using a 10-million-compound library constructed from 5,000 and 2,000 randomly selected Syn1s or Syn2s, respectively. Overall, CombiDOCK outperforms existing fragment-enrichment protocol adapted from V-SYNTHES by 7- to 20-fold in recall across all three targets, while maintaining close agreement with full docking rankings (Pearson/Spearman $R = 0.67$ to 0.89) (**Fig. S15A**, more details in **Methods**). While fragment-enrichment significantly reduces computation, it does so at the cost of losing context-dependent interactions that emerge when synthons are covalently joined, leading to poor overlap with full molecule docking results. By instead docking capped synthons only once and recombining cached poses, CombiDOCK preserves the full context of binding poses while avoiding redundant computations. For MINT-Dock, we compared it to the fragment-enrichment protocol above under matched-scale outputs for both σ_2 and VMAT2. As shown in **Fig. S15B-C**, MINT-Dock's guided search, which uses product-level docking feedback to refine its exploration, generates molecules that are not only well-scored but also more likely to satisfy the specific interaction patterns required for productive binding. **Fig. S16** further provides composition analysis that explains the increased enrichment for σ_2 after interaction-based triage.

These differences translated into prospective performance (**Fig. S15D**; see **Methods**): across σ_2 , VMAT2, and VACHT, fragment-enrichment recovered only a small fraction of the true actives identified by CombiDOCK. Even under generous conditions, where the top 6 million molecules were enumerated and docked as full compounds, fragment-enrichment still missed more than 95% of the experimental hits from the CombiDOCK campaigns. Only two σ_2 hits were recovered under the same chemical space. A similar pattern was observed for VMAT2, where fragment-enrichment identified only one hit (25%, 1/4) and no ligands with potency better than 1 μM . The VACHT campaign showed the same trend: fragment-enrichment yielded only five hits (38%, 5/13), with affinities remaining near the 1 μM range. Recall improved only when very large fractions of the library were enumerated and fully docked (**Fig. S15D** and **Table S5**). For VACHT, complete recovery of CombiDOCK actives required enumeration of more than 1 billion products, increasing the end-to-end cost to 179 cluster-days on 2,000 CPU cores, compared with 26 cluster-days for CombiDOCK.

Computational efficiency of both methods

For Enamine REAL Space (103 billion molecules), we estimated that conventional full-molecule workflows would require approximately 18,000 CPU-days and 50,000 TB of storage for conformer preparation, whereas CombiDOCK reduced this to about 1 cluster-day and 1 TB by representing the library at the synthon level (**Fig. 1A**). Average docking speed improved from 1 s per molecule for full-molecule docking (DOCK3) to 0.07 s per molecule for CombiDOCK. This corresponds to >10-fold acceleration over conventional full-molecule docking, 14-fold acceleration over AutoDock Vina GPU (1 s per molecule), and at least 300-fold acceleration over recent deep-learning co-folding methods such as Boltz-2 (20 s per molecule) and AF3 (120 s per molecule).^{103,117–119} MINT-Dock further reduced cost by docking only selected synthon combinations, typically generating >10 million docked, synthesizable molecules in 3–4 cluster days using 75–110 CPU cores. Across prospective campaigns, this corresponded to 22-fold less docking computation than the prior 490M σ_2 brute-force campaign, 64-fold less than the 1.9B VMAT2 campaign, and >80-fold less for the multi-target 1.7B VACHT campaigns. Rank-enrichment analyses against brute-force baselines showed strong concentration of well-scored molecules, including up to 660-fold rank enrichment for σ_2 , up to 130-fold enrichment for VMAT2, up to 200-fold enrichment for the VACHT selectivity task, and up to 700-fold enrichment for the VACHT polypharmacology task (**Fig. S17**), while maintaining or improving ligand-discovery outcomes relative to these prior billion-scale screens. MINT-Dock's practical advantage extends beyond reduced docking time to a major reduction in storage burden: by replacing full-molecule conformations with synthon conformations, storage requirements decrease from approximately 240 TB (490M molecules for σ_2) or 930 TB (1.9B molecules for VMAT2 and VACHT) to only 1 TB for synthon conformers, a reduction of up to nearly 1,000-fold.

The recently published DrugCLIP is able to score over trillions of ligand-protein interactions per cluster day using 8 GPUs, which has enabled genome-wide docking, discovering potent binders prospectively for three biological targets.⁶ However, extrapolating DrugCLIP's reported scaling (see details in **Methods**) suggests that screening 100 billion molecules would require 200 cluster-days for retrieval plus 116 cluster-days for conformer generation, for a total of at least 316 cluster-days under optimistic assumptions. By contrast, CombiDOCK required 1 cluster-day for synthon preparation and 30 cluster-days for screening. DrugCLIP also failed to recover the canonical ligand-binding sites for σ_2 , VMAT2, and VACHT, and its top hits scored substantially worse than CombiDOCK/MINT-Dock candidate hits after re-docking (**Fig. S11C,D**).

To test whether the improved prospective outcomes from larger libraries could be explained by simple differences in ligand size or selection depth across campaigns, we analyzed ligand heavy-atom counts and performed matched comparisons under overlapping ranking constraints. Larger libraries continued to yield higher hit rates for σ_2 , VMAT2, and VACHT across increasing heavy-atom cutoffs and under shared-ranking analysis, whereas no correlation was observed between ligand size and binding affinity among active compounds (**Fig. S18**). These results indicate that the improved performance was not driven by larger ligands or selection depth, but instead reflects improved ligand discovery outcomes in larger chemical spaces, enabled by our newly developed methods.

Discussion

Altogether, we present over 250 billion scored complexes (56 billion for σ_2 , 103 billion for VMAT2, and 103 billion for VACHT). This is almost 40-fold larger than that reported in lsd.docking.org (6.3 billion),¹²⁰ a dedicated database for large-scale docking. Additionally, we also present experimental activities for 1300 molecules across the three targets. This experimental dataset includes both active and inactive compounds, representing successful docking predictions as well as false positives. Moreover, we see great value in these datasets for benchmarking and training

ML models for virtual screening,^{121,122} and they may further support advances in computational protein-binder design.^{123,124} Nevertheless, not all ML approaches share equal promise for scaling to screening billions of molecules. Recent benchmarking on co-folding methods such as AF3¹⁰³ and Boltz2,¹¹⁹ reveals that these models perform poorly on out-of-distribution targets relative to their training set,¹²⁵ a result that is borne out in our own comparisons of AF3 with CombiDOCK. CombiDOCK's strong performance in both retrospective benchmarks and prospective docking screens highlights the advantages of this physics-based method, which offers superior scalability and generalizability for prospective screening of ultra-large libraries. These advantages also enable CombiDOCK to serve a dual role as a docking engine to quickly generate large datasets of docking scores for benchmarking and training ML models. Looking forward, as new reactions and diverse building blocks are discovered, the modular nature of make-on-demand libraries will support continued combinatorial growth. In this context, and in line with the "bigger is better" paradigm, docking at the trillion-molecule scale becomes increasingly feasible through both methods. CombiDOCK can complete a one-trillion-molecule screen in just one year on 2,000 CPU cores, with further acceleration achievable through cloud resources such as AWS.¹²⁶ These computational speedups extend beyond overcoming the limitations of brute-force molecular docking. Even for approximation strategies scaled to 1 trillion, such as ML-based or fragment-based docking, top 10% full-molecule docking remains necessary for optimal performance, making 100-billion-scale redocking computationally prohibitive. We believe CombiDOCK can serve as a practical and scalable docking engine to enable efficient docking of such enriched subsets of chemical space. This naturally leads to MINT-Dock, which leverages CombiDOCK's pre-computed synthon conformations to perform on-the-fly sampling and generation during docking while maintaining its computational speedups, and further extends this capability to multi-target design objectives that would otherwise require multiple exhaustive screens. MINT-Dock further strikes a practical balance among existing generative approaches by combining four goals that are rarely achieved simultaneously: (i) 3D, structure-based discovery through physics-based docking in the binding site, (ii) strict synthesizability through combinatorial chemistry, (iii) computational efficiency through search that evaluates only a small, high-value subset of an ultra-large library and (iv) practical throughput, enabling generation of over 100 million docked, synthesizable designs from trillion-scale chemical library with 2-3 days using 2,000 CPU cores.

Several caveats should be noted. First, three-component reactions were handled here as pseudo-two-component reactions by pre-reacting dependent synthon sets; this worked well at the current scale and remains practical at the trillion-molecule scale, although a more general hierarchical treatment of three-component assembly will become important as libraries continue to grow beyond 10 trillion molecules. Second, our multi-target studies used scalar design functions, which worked well here because the objectives were naturally commensurate, namely docking scores produced by the same docking method, making additive or monotonic combinations reasonable. Although we did not rely on explicit multi-objective MCTS for the multi-target studies presented here, MINT-Dock includes a multi-objective MCTS mode for settings in which maintaining the Pareto front is necessary.

While these limitations exist, they do not diminish the key conclusions drawn from this study. CombiDOCK and MINT-Dock together define a unified framework for navigating the trillion-molecule frontier of drug discovery. For single-target campaigns, the two methods can be deployed in a staged workflow: MINT-Dock first provides efficient directed exploration to rapidly identify promising chemotypes with modest computational resources, followed by CombiDOCK's exhaustive screen of the full combinatorial space to maximize hit rates and ligand potency, as demonstrated by CombiDOCK's consistently higher hit rates and best-in-series affinities across all three targets in this study. For multi-target objectives, including selectivity, polypharmacology, and multi-property design, MINT-Dock's ability to embed explicit design functions into the search

process makes it the method of choice, enabling discovery modes that are otherwise computationally out of reach even at the billion-molecule scale, let alone at the trillion-molecule frontier.

Figures

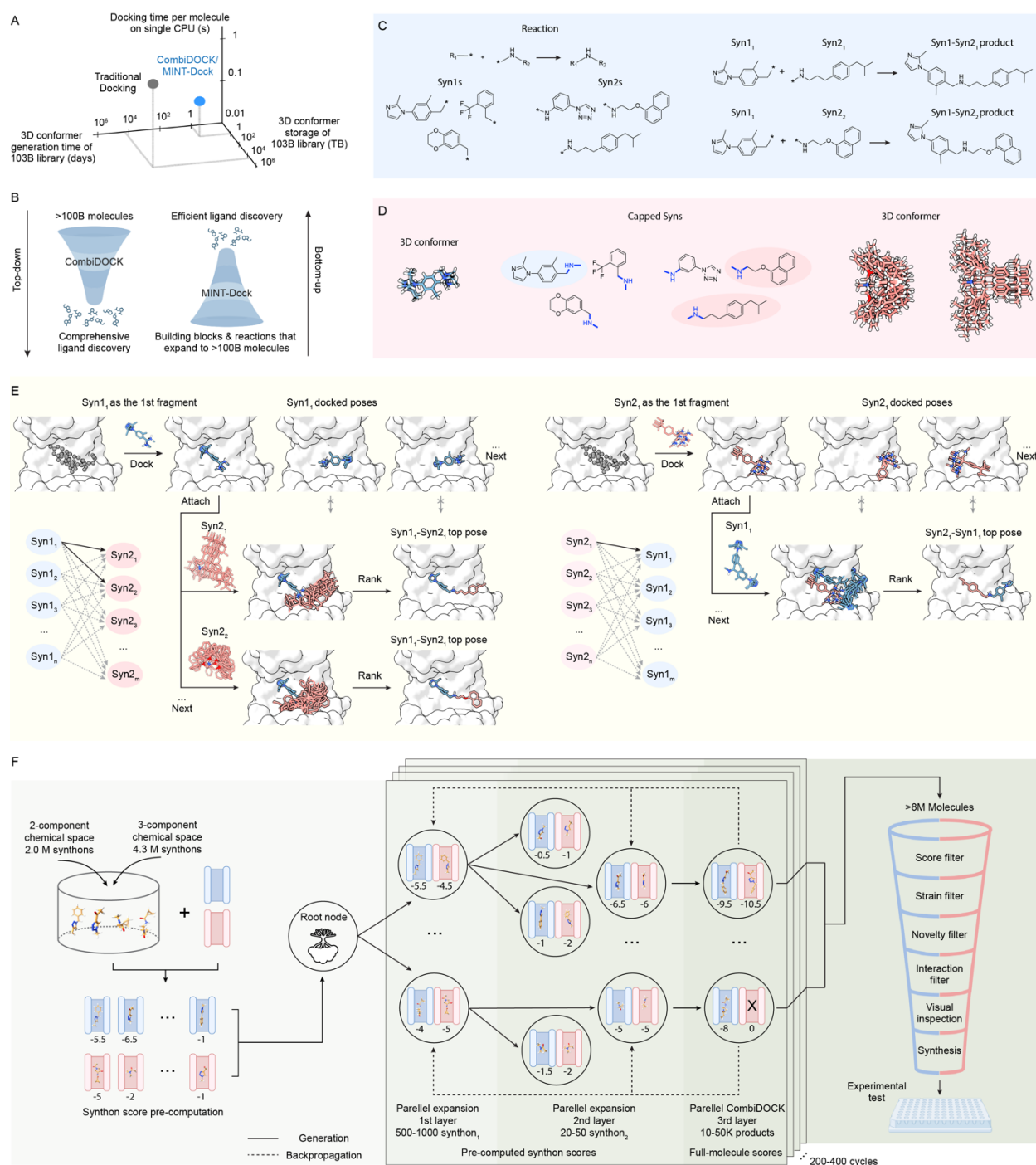


Figure 1. CombiDOCK and MINT-Dock enable exhaustive and guided navigation of ultra-large combinatorial chemical space. (A) Comparison of the CombiDOCK/MINT-Dock workflow with the traditional full-molecule docking workflow, represented here by UCSF DOCK, across three major computational dimensions: time required for 3D conformer generation, storage required for the library, and docking time per molecule. CombiDOCK/MINT-Dock improves efficiency across all three dimensions relative to conventional full-molecule docking. Data are available in **Table S1**. (B) Conceptual relationship between the two methods. CombiDOCK performs top-down, exhaustive screening of ultra-large chemical

libraries, whereas MINT-Dock performs bottom-up, guided exploration of the same chemical space for efficient ligand discovery. (C) The REAL Space virtual library was constructed combinatorially from 304 chemical reactions and 2.3 million synthons. One representative reaction type is shown, illustrating how synthons (Syn1s and Syn2s) can be combined to generate diverse products. (D) Examples of capped synthons and the hierarchical organization of their precomputed 3D conformers. (E) Overview of the CombiDOCK workflow, showing how docking poses of the first synthon are reused across multiple product combinations to accelerate virtual screening. Nonpolar hydrogens are hidden in this panel for clarity. (F) The Overview of the MINT-Dock workflow, shown here for a multi-target optimization task. Precomputed synthon scores guide Monte Carlo Tree Search through synthon space, while full-molecule scores from the CombiDOCK engine are used for product evaluation and backpropagation, enabling efficient sampling of a small, high-value subset of ultra-large chemical space.

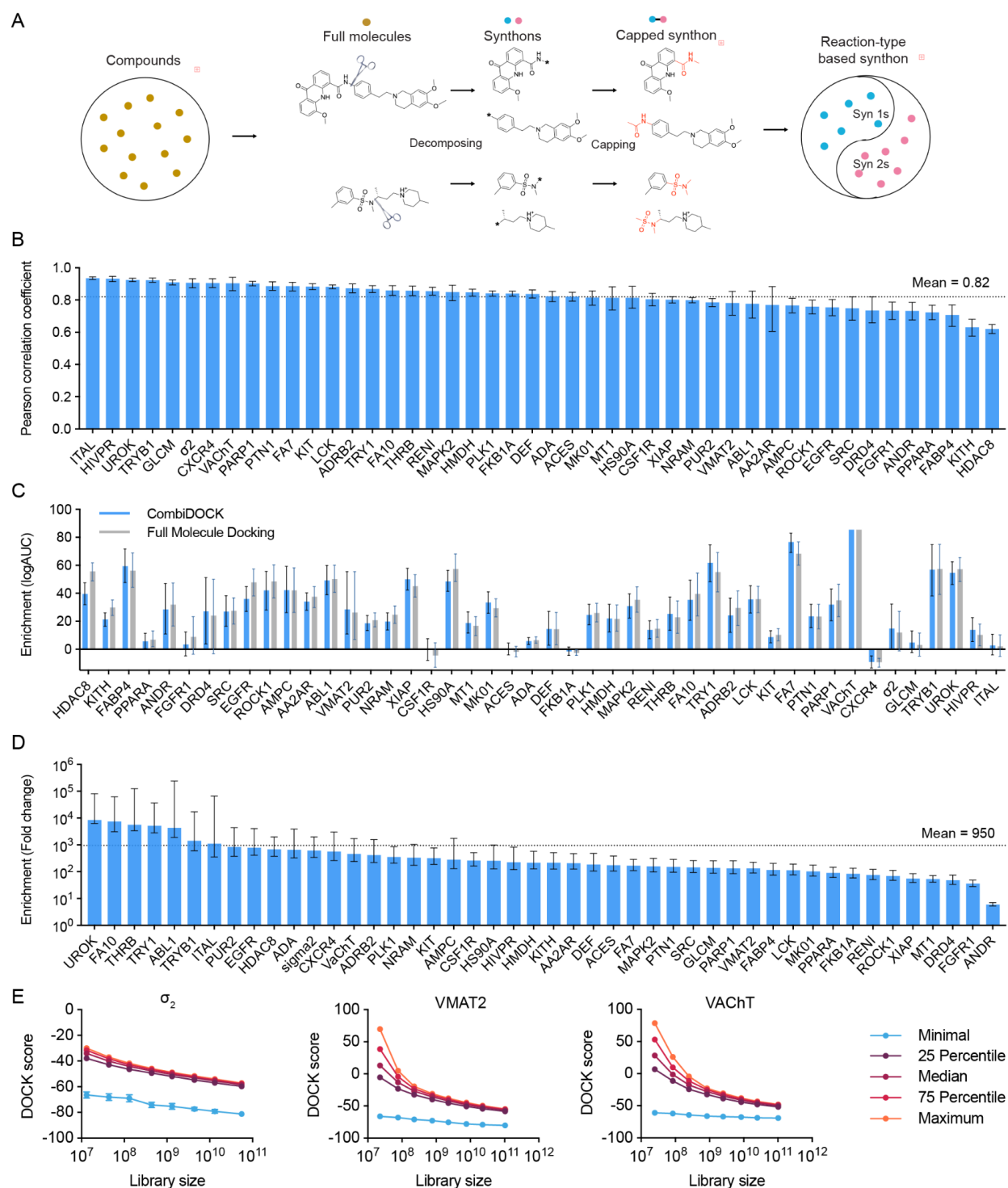


Figure 2. CombiDOCK approach to navigate Enamine REAL Space. (A) Compounds from pre-assembled libraries were decomposed into pairwise synthons, enabling combinatorial reassembly during docking. (B-C). Across 46 protein targets — including 43 from the DUDE-Z benchmark as well as the σ_2 receptor (TMEM97), VMAT2 (SLC18A2), and VACHT (SLC18A3) — CombiDOCK achieved performance comparable to full-molecule docking, as assessed by B, Pearson correlation of docking scores and C, logAUC-based early enrichment. Error bars represent 90% confidence intervals, estimated by bootstrapping (1,000 iterations with replacement). (D) Summary plot of the adjustment enrichment ratio

between molecules generated by MINT-Dock and molecules randomly sampled at a designated score threshold for all 46 targets, where the random library was docked by CombiDOCK. Error bars represent 90% confidence intervals, estimated by bootstrapping (1,000 iterations with replacement). (E) Docking score improves with increasing library size, illustrated for σ_2 (56 billion), VMAT2 (103 billion), and VACHT (103 billion).

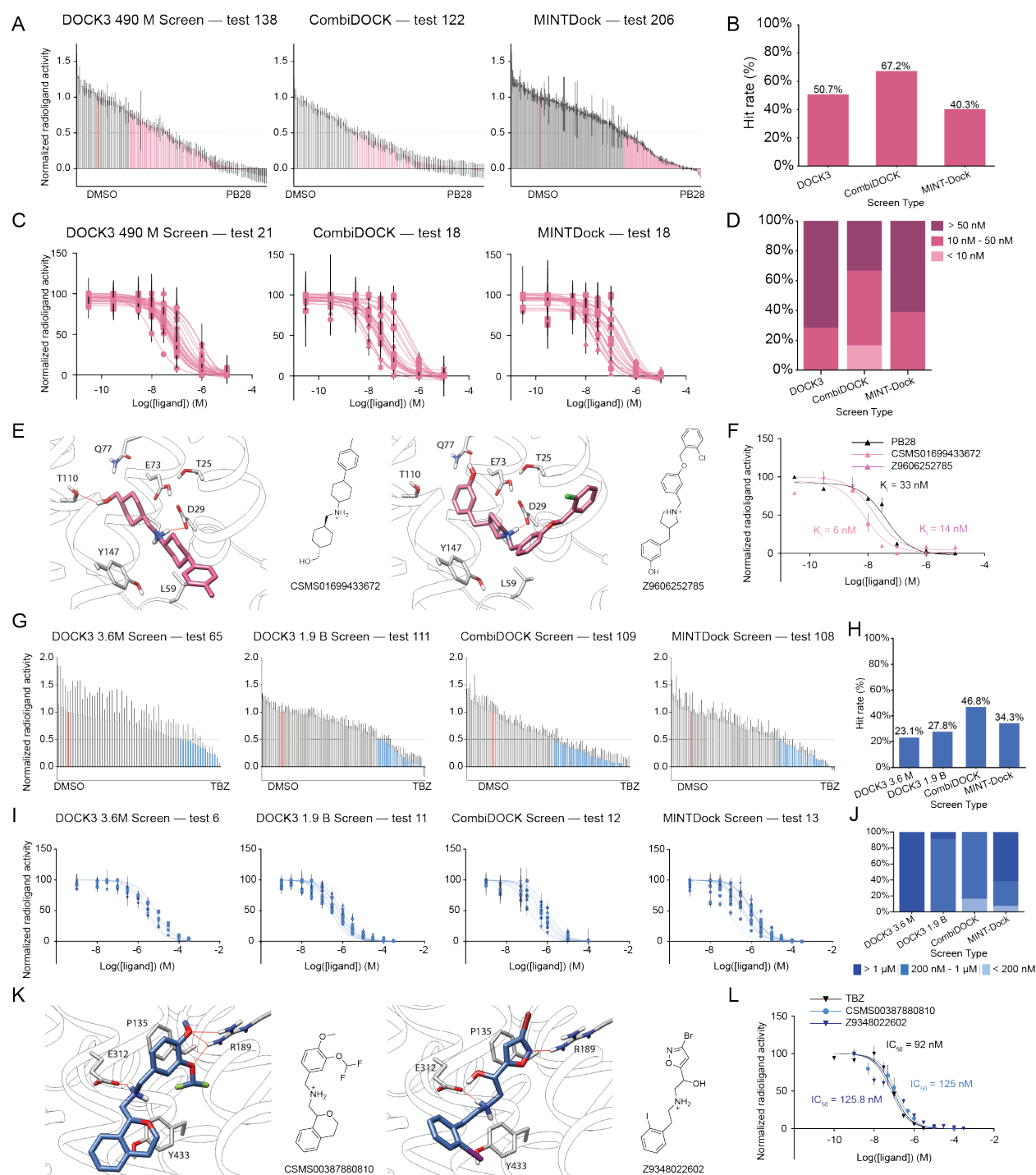


Figure 3. Summary of virtual screening results against σ_2 and VMAT2. (A) In a [3 H]-DTG binding assay, using > 50% [3 H]-DTG displacement as a hit threshold, out of 138 tested compounds from the 490 million screen against σ_2 , 70 hits were identified (left) representing a hit rate of 51%. From the 122 tested compounds from the 56 billion screen 82 hits were identified (middle), representing a hit rate of 67%. From the 206 compounds tested from the MINT-Dock 103 billion screen, 83 hits were identified (right), representing a hit rate of 40.3%. (B) Hit rates for 490 million, CombiDOCK 56 billion, and MINT-Dock 103 billion sized library screens. (C) Dose-response curves of 21 re-tested compounds from the previous 490

million screen (left), 18 tested compounds from the CombiDOCK screen (middle), and 18 tested compounds from the MINT-Dock screen. Curve of positive control PB28 (red curve) is in dose-response curves from the previous 490 million screen (left). (D) The potency distribution shifts towards higher potency compounds from the 490 million screen to the larger CombiDOCK and MINT-Dock screens. (E) Predicted pose and structure of the most potent compound from the CombiDOCK screen (left), and the MINT-Dock screen (right). (F) The dose-response curves of the most potent compound from the CombiDOCK screen and the MINT-Dock screen, with PB28 as the control. (G) In a [^3H]-dopamine uptake assay, using > 50% [^3H]-dopamine displacement as a hit threshold, 15 out of 65 tested compounds from the 3.6 million screen were hits representing a hit rate of 23% (top left). In the 1.9 billion screen against VMAT2, 30 out of 111 tested compounds were hits representing a hit rate of 27% (middle-left). In the 103 billion CombiDOCK screen against VMAT2, 51 out of 109 tested compounds were hits representing a hit rate of 47% (middle right). In the 103 billion MINT-Dock guided screen, 37 out of 108 tested compounds were hits representing a hit rate of 34% (right). (H) Hit rate steadily improves as screening size grows. (I) Dose-response curves of tested compounds across screening size. (J) The potency distribution shifts towards higher potency compounds as the screening size grows. (K) Predicted poses, and structures of the most potent compounds from the 103 billion screen against VMAT2 using CombiDOCK (left) and MINT-Dock (right). (L) The dose-response curves of the most potent compound from the CombiDOCK screen and the MINT-Dock screen, with TBZ as the control.

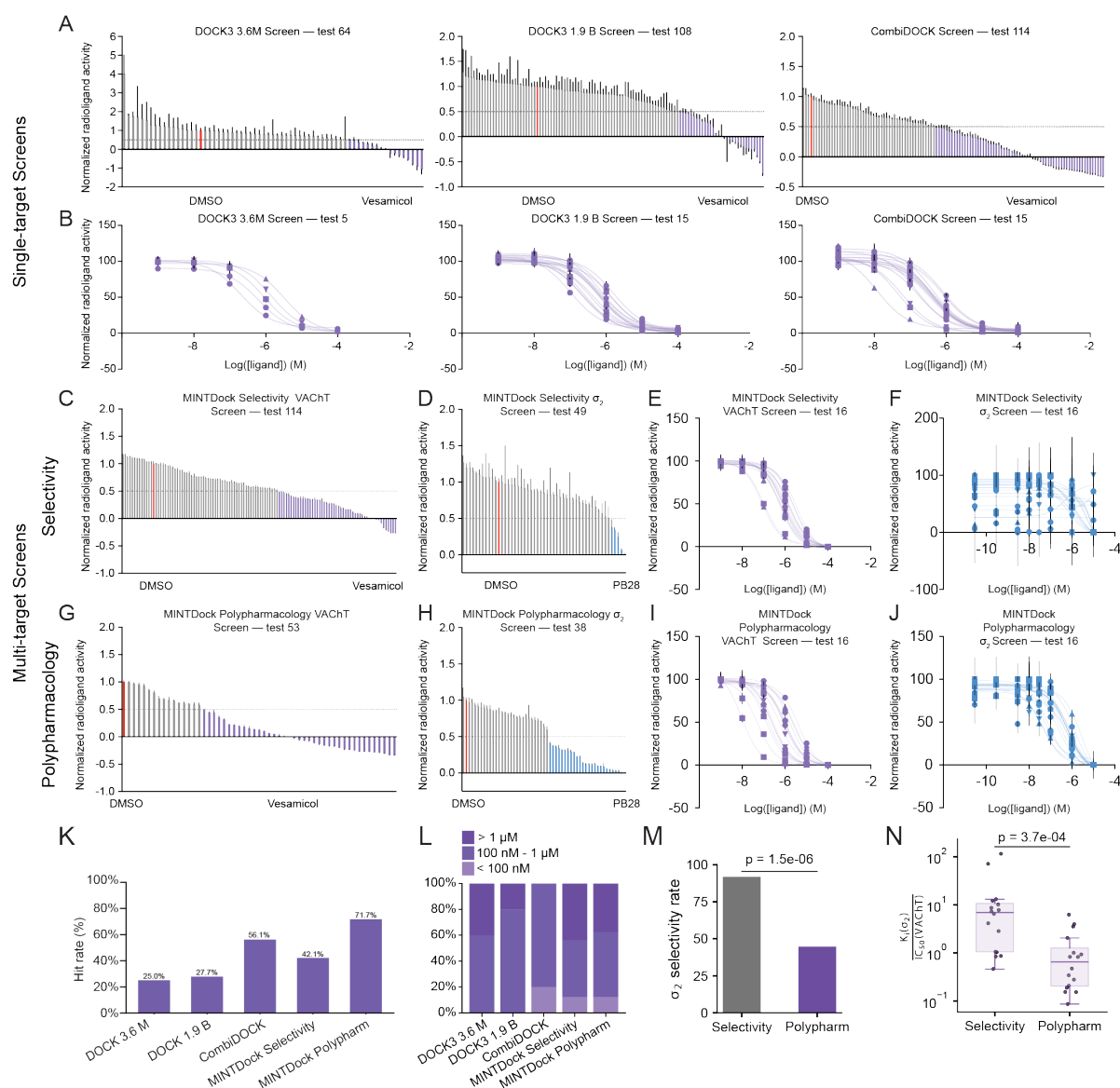


Figure 4. Summary of virtual screening results against VACHT. (A) VACHT compounds from the DOCK3 and CombiDOCK screens were tested in a [³H]-vesamicol uptake assay, using > 50% [³H]-vesamicol displacement as a hit threshold. 16 out of 64 tested compounds from the 3.6 million screen were hits, representing a hit rate of 25% (left). 15 out of 108 tested compounds from the 1.9 billion screen were hits, representing a hit rate of 28% (middle). 64 out of the 114 tested compounds were hits in the 103 billion CombiDOCK screen, representing a hit rate of 56% (right). (B) Dose-response curves of tested compounds from the DOCK3 and CombiDOCK screens. (C) 49 out of 114 tested compounds from the 103 billion MINT-Dock selectivity screen were found to be hits when tested in a [³H]-vesamicol uptake assay, representing a hit rate of 43%. (D) 45 of the 49 VACHT hits from the MINT-Dock selectivity screen displaced <50% [³H]-DTG in a [³H]-DTG binding assay. (E) VACHT dose-response curves of compounds from MINT-Dock selectivity screen. (F) σ_2 dose-response curves of compounds from MINT-Dock selectivity screen. (G) 38 of the 53 tested compounds from the MINT-Dock polypharmacology screen were found to be hits when tested in a [³H]-vesamicol uptake assay, representing a hit rate of 72%. (H) 21 of the 38 VACHT hits from the MINT-Dock polypharmacology screen displaced >50% [³H]-DTG in a [³H]-DTG binding assay. (I) VACHT dose-response curves of compounds from MINT-Dock polypharmacology screen. (J) σ_2 dose-response curves of compounds from MINT-Dock polypharmacology screen. (K) Hit rate steadily improves as library size increases. (L) The potency distribution shifts towards higher potency compounds as the

screening size grows. (M) Selectivity hit rate comparison between the two MINT-Dock multi-target campaigns. P-value was calculated by the two-sample proportion Z-test. (N) Selectivity fold comparison between the two MINT-Dock multi-target campaigns. P-value was calculated by the two-sided Mann-Whitney U test.

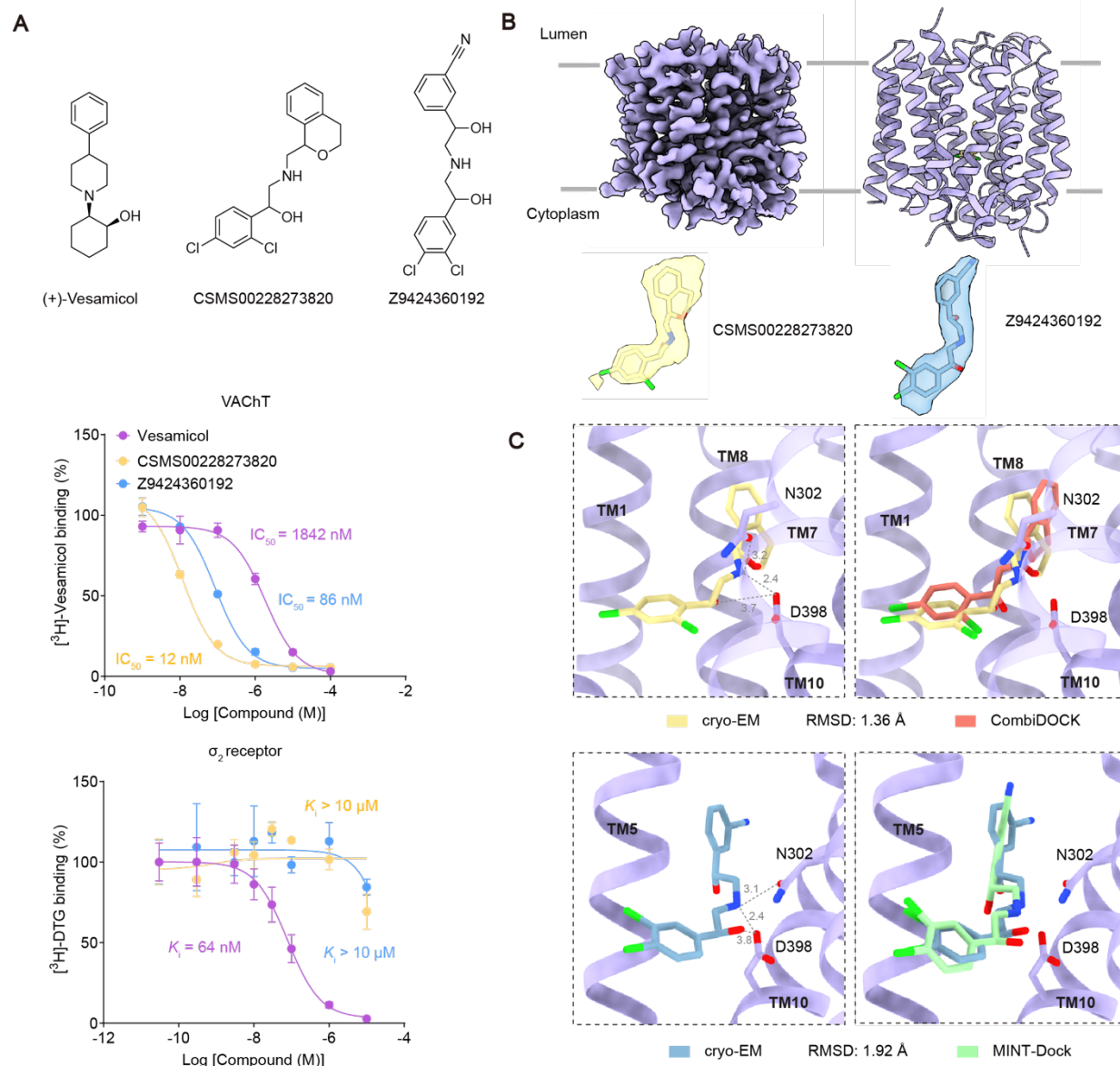


Figure 5. Structural determination of VACHT bound with ligand CSMS00228273820 and Z9424360192. (A) The 2D structure of vesamicol, CSMS00228273820 and Z9424360192, along with dose-response curves of vesamicol (blue), CSMS00228273820 (purple) and Z9424360192 (yellow) against VACHT (top) and σ_2 (bottom). (B) CryoEM density map of ligand-bound VACHT (top). The model of CSMS00228273820 and Z9424360192 built into density is shown (bottom). (C) Experimental poses of CSMS00228273820 (top) and Z9424360192 (bottom) are shown to conserve key interaction with D398. High-fidelity between the experimental pose and the CombiDOCK predicted pose of both CSMS00228273820 and Z9424360192 is observed with an RMSD < 2 Å.

Methods

Preparation of DOCKable files for synthon enumeration library

The reactions and corresponding synthons comprising the enumeration library (the version of April 2023) were provided by Enamine. This database consists of 304 chemical reactions with 2.3 million synthons, theoretically enabling the enumeration of 36 billion compounds without considering stereoisomer expansion. All the reactions in the database can be separated into two categories: two-component and three component reactions. Since the combination of synthons in three component reactions has the dependency that one synthon is the linker connecting other two synthons, we simplified them into “pseudo two component reactions” by merging the linker synthon with one of the other two synthons. Moreover, we excluded 4 chemical reactions that generate the covalent compound libraries.

At the synthon level, to ensure the retention of reaction-formed functional groups, all the synthons in the database were capped according to their corresponding reaction-formed functional groups. These capping groups also served as chemically meaningful junctions for aligning pairwise synthons and subsequently reconstructing full molecules during docking. The capped synthons were represented in SMILES strings and were subsequently used to generate hierarchical trees of precomputed conformers in DB2 format.¹²⁷ The ligand conformer preparation pipeline involved protonation state and tautomer enumeration using Jchem (v.22.22.0, <https://chemaxon.com/>), stereoisomers expansion via RDKit (v2023.09.1),¹²⁸ three-dimensional structures using Corina (v.4.4.0026, <https://mn-am.com/products/corina/>), and conformational ensemble generation with OMEGA (v.2019.10.2, <https://www.eyesopen.com/omega>).¹²⁹ Atomic partial charges and desolvation energies^{130,131} were computed as previously described.¹³²

The resulting curated library of 3D capped synthons enables the theoretical construction of 103.2 billion compounds: 5.5 billion from two component reactions and 97.7 billion from pseudo two component reactions. Applying a heavy atom count threshold of 34 for pseudo two component products reduced the final enumerated docking library to approximately 56 billion compounds. For the σ_2 receptor, we screened these 56 billion compounds, while for VMAT2 and VACHT, we screened the whole library of 103.2 billion compounds.

Combinatorial docking approach

Our combinatorial docking approach, implemented in the version 6.14 of DOCK6,⁴⁵ follows a hierarchical sampling strategy using capped synthons that sequentially assemble into final products. The process begins by docking a hierarchical tree of precomputed conformers of the *first* synthon (Syn1) into the target protein pocket, generating multiple Syn1 docking poses per orientation (this is a tunable docking parameter called *number per first search*, default value is 10). The hierarchical tree conformer of the *second* synthon (Syn2) is then grown onto these docked Syn1 poses to generate feasible docking poses for the complete molecule (governed by another tunable docking parameter called *number per second search*, default value is 5). Each resulting pose is scored using the DOCK 3.5 scoring function (ChemGrid Score),⁵⁰ and the best-scoring pose of each full molecule is retained for VS. The combinatorial docking parameters for

CombiDOCK are illustrated in **Table S6**, and the technical details were presented in the later sections.

Docking models and parameters for large-scale docking

The σ_2 receptor bound to cholesterol (PDB ID: 7MFI),²⁷ the VMAT2 receptor bound to tetrabenazine (PDB ID: 8T69),¹¹³ and the VACHT receptor bound to vesamicol (PDB ID: 8ZMR)¹⁰¹ were used as the 3D structural input for the σ_2 , VMAT2, and VACHT docking campaigns, respectively. In the σ_2 receptor docking model, the structure was prepared as previously described. In the VMAT2 docking model, 45 matching spheres were calculated around the tetrabenazine ligand atoms, and AMBER united atom charges were assigned to the receptor structure. The volume of the low dielectric and the desolvation volume was extended out 1.2 Å and 0.1 Å, respectively, from the surface of protein using spheres calculated by SPHGEN. Energy grids were pre-generated with AMBER force fields using CHEMGRID for van der Waals potential,¹³³ QNIFFT¹³⁴ for Poisson-Boltzmann-based electrostatic potentials and SOLVMAP¹³¹ for ligand desolvation. This docking model was evaluated for its ability to enrich 5 known VMAT2 ligands (DTBZ, TBZOH, tetrabenazine, valbenazine, deutetabenazine) over 250 property-matched decoys. Decoys are unlikely to bind to the receptor because despite their similar physical properties to known ligands, they are topologically dissimilar. The docking model described above was able to achieve a logAUC of 2.18 and to recover the crystal pose of tetrabenazine with RMSD values of 0.39 Å. We also constructed an “extrema” set of 28,086 molecules using the DUDE-Z webserver (<http://tldr.docking.org>)⁸⁶ to ensure that molecules with extreme physical properties were not enriched. This docking model enriched close to 94% monocations among the top 1,000 ranking molecules. Since the docking grids were originally generated using DOCK 3.7 Blastermaster for DOCK3, we converted them to DOCK6 readable grids using the method described by Balias and colleagues¹³⁵ and using their script (available here: https://github.com/tbalias/teb_docking_test_sets.git).⁸⁴ The VACHT docking model was similarly prepared as described above. Matching sphere perturbation was skipped. The low dielectric and desolvation volumes were extended out 1.7 Å and 0.3 Å, respectively. Energy grids were pre-generated using CHEMGRID,¹³³ QNIFFT,¹³⁴ and SOLVMAP.¹³¹ Using retrospective docking, the docking model was evaluated on how well it enriched four known VACHT ligands (vesamicol, ChEMBL1089205, ChEMBL20730, ChEMBL3587324) over 200 property-matched decoys. The model achieved a high logAUC of 64.26, and successfully recapitulated the crystal pose of vesamicol with an RMSD value of 1.61 Å. An “extrema” set of 79,815 molecules was generated and docked, revealing over 92% of the top 1,000 ranking molecules to be monocations. Finally, the scoring grids from the docking model, generated using DOCK3.7 Blastermaster, were converted into a DOCK6 readable format using the method/script described above.

Because CombiDOCK evaluates full molecules through on-the-fly assembly guided by pre-docked synthon ensembles, performance depends on two sampling parameters: the number per first search (how many top poses of the first synthon per orientation are retained for potential extension to partner synthons, averaging 2,500 orientations per molecule) and the number per second search (multiplied by the number per first search to determine how many top poses of the fully combined molecule are passed into the minimizer). On retrospective σ_2 and VMAT2 datasets, docking score distributions and Pearson correlations with full docking rose sharply at small values

of number per first search and then plateaued (**Fig. S6A-C**). Similar trends were observed as number per second search increased (**Fig. S6D-F**). Specifically, performance converged after sampling the top 10 first-synthon poses per orientation and the top 10 combined molecules per minimizer. Since larger parameter values proportionally increase docking runtime, for the three prospective campaigns below, we therefore set the number per first search to 10 and the number per second search to 5, which together capture near-maximal agreement with full docking (Pearson $R = 0.86$ for σ_2 and 0.74 for VMAT2) while minimizing computational cost.

Benchmarking datasets for evaluating CombiDOCK performance

To evaluate the performance of CombiDOCK, we first benchmarked it against conventional full-molecule docking using three curated ligand-decoy datasets: σ_2 , VMAT2, and VACHT. The σ_2 dataset comprised 10 known ligands and 550 property-matched decoys derived from a previous study. The VMAT2 dataset consisted of 5 actives and 250 matched decoys, as described above in the VMAT2 docking model evaluation. The VACHT dataset contained 4 active compounds and 200 decoys. After stereoisomer enumeration and successful 3D conformer generation, all molecules were stored in DB2 format for full-molecule docking. For CombiDOCK benchmarking, each molecule was decomposed into pairwise capped synthons according to common reaction types, such as ether, amide, or sulfonamide linkages. These capped groups served as chemically meaningful junctions for aligning synthons and reconstructing complete molecules during the docking process. To ensure conformer consistency, 3D conformers of full molecules were used to derive the conformers of their corresponding synthons, which were also stored in DB2 format for combinatorial docking.

It should be noted that 57 molecules in the σ_2 dataset—including 3 actives and 54 decoys—could not be feasibly decomposed into synthons. For these compounds, full molecule docking scores were directly incorporated into the CombiDOCK benchmarking analysis. Additionally, we evaluated the impact of the key docking parameters on docking performance. As shown in **Fig. S6C**, increasing the *number per first search* parameter improved the correlation between CombiDOCK and full-molecule docking scores. To balance computational cost and accuracy in large-scale docking campaigns, this parameter was set to a default value of 10. Similarly, we explored the *number per second search* parameter and set it to a default value of 5 (**Fig. S6F**).

Moreover, to assess generalizability beyond these three targets, we extended the evaluation to 43 additional protein targets from the DUDE-Z benchmark,⁸⁶ each comprising actives and property-matched decoys at a 1:50 ratio. For consistency, 3D conformers of the full molecules were generated and subsequently decomposed into pairwise capped synthons for combinatorial docking.

MINT-Dock implementation

The Enamine REAL Space used in this study is composed of 2.3 million synthons and encompasses nearly 103 billion small molecules. To fit three-component reactions to the CombiDOCK format, we converted three-component reactions to pseudo-two component reactions, which raises the total number of synthons to around 6.3 million. Since one building

block can potentially react in multiple reactions, we mapped each unique (building block SMILES, reaction_id) pair to a unique identifier. In the following, we will use the term “synthon” to refer to each (building block SMILES, reaction_id) pair with its own identifier. We then cap each synthon according to each reaction it will participate in. Briefly, we replaced the terminal dummy atom with a connecting functional group corresponding to each reaction. We then used DOCK6 to compute the docking score of each synthon. If a synthon has multiple conformers with different docking scores, we took the minimum (most favorable) score. We capped the minimum synthon score to -10 and set a score of 10 for all synthon scores worse than -10. Three-component reactions were converted to two-component reactions, and the resulting synthons were followed by the same preprocessing procedure.

In the actual MCTS, each node is scored by a scoring function, which is a function of the node itself and the rollout number N , similar to the Synthemol implementation:⁷⁴ $S(\text{node}, N) = \frac{Q(\text{node}, N) + c_2 P(\text{node}) U(\text{node}, N)}{c_1}$, where for all non-product layer, $Q(\text{node}, N) = \sum_{\text{cycles}} q(\text{node}, N) = \sum_{\text{cycles}} \sum_{\text{child nodes}} q(\text{child node}, N)$, which represents the exploitation score. If node is the product layer, then $q(\text{node}, N) = f(S_{\text{node}}, C_{\text{node}}) = S_{\text{node}} - c_3 |C_{\text{node}} - C_{\text{empirical}}|$. $S_{\text{node}}, C_{\text{node}}$ are the docking score and the formal charge of the product contained in that node, and $C_{\text{empirical}}$ is the empirical charge for a good molecule that could reasonably reach the target in vivo and bind with the target. c_3 is a penalty hyperparameter that tunes how much penalty is placed on deviation in the product charge. $U(\text{node}, N) = \frac{\sqrt{1+n_{\text{total}}}}{1+n_{\text{node}}}$, where n_{node} is how many times the node has been visited and $n_{\text{total}} = \sum_{\text{sibling}} n_{\text{sibling}}$. This represents the exploration score followed from a prior MCTS scoring function.¹³⁶ $P(\text{node})$ is the average of precomputed scores of synthons contained in that node, and c_2 is an explore weight that tunes how much emphasis is placed on exploration versus exploitation. Finally, c_1 is a normalization hyperparameter that scales $S(\text{node}, N)$. Since the DOCK6 score is retrieved at 6 decimal points precision, we also round $S(\text{node}, N)$ to 6 decimal points.

We set the first layer nodes to the synthons that successfully docked in the target with a docking score below -10. In the first layer, we adopted a rolling visit strategy, where the tree first randomly selects a fraction of all available nodes for further expansion, which allows for more global explorations and provides speed-up for the tree search process. The tree then keeps the nodes that are estimated to have enough synthon₂ to expand in the second layer and selects N_1 (=500 or 1000) nodes with highest scores to continue in the second layer. In the second layer, for each selected first layer node, the tree searches for nodes that have synthon₂ entries compatible with the synthon₁ in that node, excludes nodes that have already been visited, and selects at most N_2 (=50 for two-component reactions, =20 for three-component reactions) nodes with highest scores to continue to a single product layer, which then proceeds to parallel CombiDOCK calculation of all $N_1 \times N_2$ synthon combinations, where N_1 Slurm jobs are submitted, each completing CombiDOCK calculation of one synthon₁ with at most N_2 synthon₂. In practice, each round usually generates less than $N_1 \times N_2$ synthon combinations due to insufficient number of expandable synthon₂ from a portion of selected reaction groups. After DOCK6 score has been computed for all products, the best score for each unique synthon combination, considering all possible conformations, is backpropagated to each parent layer to update n_{node} and $q(\text{node}, N)$. We set

$c_{\text{empirical}} = +1$ and a high $c_3 = 50$ to penalize di-cations with favorable DOCK scores for all prospective screening campaigns.

For MINT-Dock for two targets, the DOCK6 scores of each synthon for each target are precomputed, and the score function becomes $S(\text{node}, N) = \frac{Q(\text{node}, N) + c_2 g\{P_{\text{target1}}(\text{node}), P_{\text{target2}}(\text{node})\} U(\text{node}, N)}{c_1}$ and for all non-product layer, $Q(\text{node}, N) = \sum_{\text{cycles}} q(\text{node}, N) = \sum_{\text{cycles}} \sum_{\text{child nodes}} g\{q_{\text{target1}}(\text{child node}, N), q_{\text{target2}}(\text{child node}, N)\}$. The expression of P and q for each target is the same in the single target case. g is a function that can be tuned according to the purpose of MCTS. For poly-pharmacology assays, we set $g(x_1, x_2) = \frac{x_1 + x_2}{2}$ and for selectivity assays where we want the selected compounds to favor target₁ over target₂, we set $g(x_1, x_2) = x_1 - \frac{x_2}{2}$ to balance between favorable docking scores for target₁ and large docking score difference between target₁ and target₂. For poly-pharmacology screens, we also differentiate between the two targets by setting a primary target (target₁) and an auxiliary target (target₂). For compounds that are unable to be docked in the pocket of the primary target, their auxiliary target score is also set to a default low value.

Since MINT-Dock utilized a dynamic Slurm scheduling procedure to complete parallel docking, we approximated the average number of cores utilized for one complete MINT-Dock screen by the total DOCK time divided by the sum of maximum DOCK time across all Slurm jobs submitted per cycle. We summed the average number of utilized cores for two-component and three-component runs to get a total average number of utilized cores for single-target screens.

MINT-Dock benchmarking

For this benchmarking, we compared the performance of MINT-Dock sampled molecules with randomly sampled molecules from an Enamine make-on-demand library of 141 reactions (131 two-component, 10 three-component).

For each target, we computed the dock score of each capped synthon for setting the initial heuristic strategy for MINT-Dock search with full molecule DOCK6 HDB docking.⁸⁴ We capped the minimum synthon score to -10 and set a score of 10 for all synthon scores worse than -10. Here, we do not impose any charge penalty. We ran 60 cycles (1.5 million rollouts) for two-component reactions and 10 cycles (0.1 million rollouts) for three-component reactions with an explore weight of 0.001, hence constraining the theoretical maximum number of the generated compounds to roughly half of the random subset's size. Relevant parameters are included in **Table S7**. For each target, MINT-Dock sampled between 3.2 to 9.0 million stereogenic molecules and eventually generated between 1.1 to 1.5 million compounds.

To procure the docking scores of random molecules, we curated 2.9 million unique compounds identified by synthon combinations from these 141 reactions totaling 5.3 million stereogenic molecules from full molecule db2 generation for DOCK6 HDB and 7.4 million stereogenic molecules from synthon db2 generation for CombiDOCK, while keeping the number of unique synthon combinations contributing to each reaction proportional to the full library. We then docked

these molecules with either CombiDOCK or DOCK6 HDB. The adjusted enrichment ratio was calculated by the formula $\frac{\sum_{i \in \text{MINT-Dock}} (s_0 - s_i)}{\sum_{i \in \text{random}} (s_0 - s_i)} \cdot \frac{n_{\text{random}}}{n_{\text{MINT-Dock}}}$, where s_0 is the chosen dock score threshold, s_i is the best dock score for compound i , and $n_{\text{MINT-Dock}}$ and n_{random} are the number of synthon pairs generated by MINT-Dock and by random subset respectively. For gross evaluation of MINT-Dock enrichment, we defined the score threshold for each target as $\max(s_{\text{MINT-Dock}}, s_{\text{random}})$, where $s_{\text{MINT-Dock}}$ is the docking score of the 1,001th ranked compound from MINT-Dock, and s_{random} is the docking score of the 2nd ranked compound from the random sample. To estimate the confidence interval, we performed bootstrapping with replacement on the MINT-Dock molecule scores and random molecule scores respectively. Namely, for each bootstrapping iteration, we calculate the adjusted enrichment using the pair of bootstrapped MINT-Dock sample and bootstrapped random sample. We repeated this process 1,000 times to gather an overall distribution to estimate the confidence interval.

MINT-Dock backpropagation comparison

We tested whether backpropagation improves MINT-Dock's ability to generate favorable molecules over successive search cycles. Measuring affinity improvement directly would be impractical, so we used docking score as a tractable proxy for search progress, since docking score correlates with experimental hit rate in three systems (AmpC beta-lactamase, D4 dopamine and σ_2 receptors) in prior studies⁹¹ and remains the primary selection criterion in docking campaigns. We compared the number of good compounds between MINT-Dock with backpropagation implemented and MINT-Dock without backpropagation implemented, where in each rollout, the tree search is based on precomputed synthon scores only, on σ_2 and VMAT2 receptors and on two multi-target objectives (**Fig. S1**). From the final pool of generated molecules, we compared the number of generated compounds with docking scores better than three given score thresholds respectively. The most lenient score threshold was chosen as -45 kcal/mol on the basis that this is the docking score at which the hit rate versus docking score curve for σ_2 starts to show gradient in a previous large-scale virtual screening campaign.²⁷ The most stringent threshold was chosen as -56.5 kcal/mol and -53.5 kcal/mol for σ_2 and VMAT2 respectively based on the docking score distribution of the molecules from a 2 billion monocation library screen with CombiDOCK, where -56.5 is roughly the lowest score of the top 300,000 ranked compounds for the σ_2 docking campaign, and -53.5 is roughly the lowest score of the top 1 million ranked compounds for the VMAT2 docking campaign. For multi-target comparison, the criterion for the number of good compounds was VACHT docking score below -45, σ_2 docking score above -40 for selectivity, or both targets docking score below -45 for polypharmacology. Backpropagation consistently increased the yield of well-scoring molecules across multiple score cutoffs, and the advantage became more pronounced as the number of cycles increased, reaching 1.2 to 2.2-fold at the most stringent thresholds (**Fig. S2**). For selectivity (VACHT selective over σ_2), we considered a compound well-scored if its VACHT docking score was < -45 kcal/mol and its σ_2 docking score was > -40 kcal/mol. For polypharmacology, we considered a compound well-scored if its docking score was < -45 kcal/mol against both VACHT and σ_2 . Under these definitions, backpropagation improved screening results for well-scored two-component compounds to 2.4-fold in both selectivity and polypharmacology campaigns (**Fig. S2**). Notably, backpropagation produced a smaller gain for σ_2 than for VMAT2, even though σ_2 yielded more well-scoring

compounds overall. We hypothesized that this reflects how well synthon-level docking approximates full-molecule docking for each system. Consistent with this, the root mean square deviation (RMSD) between synthon and full-molecule poses is generally lower for σ_2 than for VMAT2 (**Fig. S3A–B**), indicating that precomputed synthon scores are a more reliable proxy for full-molecule scores in the σ_2 campaign. This difference is especially pronounced for first-layer synthons (**Fig. S3C–D**), where the tree is widest and early selection is strongest, which suggests that the σ_2 performance relies more on synthon priors and reduces the incremental benefit of correction by backpropagation. In contrast, for VMAT2, larger synthon-to-product pose discrepancies make product-level docking feedback more informative, leading to a stronger backpropagation effect.

MINT-Dock synthon versus full-molecule RMSD calculation

For both σ_2 and VMAT2, we took the top 300,000 ranked compounds out of their respective 10 million rollout MINT-Dock runs for two-component reactions after filtering for formal charges greater than +1.1 and removing compounds with duplicate SMILES. We then separated each full molecule into their synthon₁ and synthon₂ part poses using Chimera 1.18,¹³⁷ maintaining the full structural information of the non-cap regions of the synthons. We calculated the RMSD between the heavy atom of each synthon component of the full molecule pose and those of the original best scoring capped synthon pose and then take either the first layer synthon RMSD or the minimum of the two RMSDs, one for each synthon component, for evaluating the distribution.

MINT-Dock comparison with ML-MCTS

To assess whether coupling MCTS to physics-based docking improves generation relative to an ML surrogate, we compared MINT-Dock to an ML-guided MCTS baseline (“ML-MCTS”, adapted from Synthemol⁷⁴). For both σ_2 and VMAT2, MINT-Dock produced substantially more well-scored molecules exceeding any given score threshold, yielding higher enrichment across thresholds (**Fig. S4**). For VMAT2, we docked approximately 10 million randomly selected molecules from ZINC with DOCK6, 590,011 molecules had scores below -10, which we used to train a Chemprop model for predicting the scores of full molecules.¹³⁸ For σ_2 , we docked approximately 2.5 million randomly selected molecules from ZINC with DOCK6 and selected 600,000 molecules that had scores below -10 to train a Chemprop model for predicting the scores of full molecules. We used these models to predict the scores for each unique SMILES to use to populate each synthon node. We adapted the Synthemol architecture so that it generates 10 molecules per cycle rather than 1 molecule per rollout to speed up the generation process. We generated 979,991 molecules for σ_2 and 981,684 molecules for VMAT2 with 1 million rollouts on the two-component space with ML-MCTS. We generated 727,545 molecules for σ_2 and 849,043 molecules for VMAT2 with 1 million rollouts on the two-component space with MINT-Dock. We then used CombiDOCK to calculate the actual scores of these molecules and compare between the two methods. The adjusted enrichment ratio was calculated by the formula presented in comparison with random docking.

MINT-Dock-based screening campaigns

For all prospective screening campaigns, we accounted for the fact that the desired small molecule binders will need to permeate through cell membranes and cannot have high charges by assigning any molecule with a charge greater than +1 a penalty that is an affine function of its charge during backpropagation as described previously.

For σ_2 , we ran 400 cycles (10 million rollouts) on the two-component space and 200 cycles (2 million rollouts) on the three-component space, which sampled 120 million stereoisomers in total and generated 8.3 million and 2.0 million compounds respectively. For VMAT2, we ran 600 cycles (15 million rollouts) on the two-component space and 200 cycles (2 million rollouts) on the three-component space while filtering for high strain energy (total torsion strain energy of less than 8 units, maximum strain energy per torsion angle of less than 1.5 units) during docking to prevent the generation of many strained poses in the narrow pocket of VMAT2, which sampled 58 million stereoisomers in total and generated 13.6 million two-component and 2.0 million three-component compounds respectively. For VACHT versus σ_2 selectivity assay, we ran 300 cycles (15 million rollouts) on the two-component space, which sampled 33 million stereoisomers and generated 12.9 million compounds. We also ran 100 cycles (2 million rollouts) on the three-component space for the purpose of backpropagation comparison only. For VACHT plus σ_2 poly-pharmacology assay, we ran 400 cycles (10 million rollouts) on the two-component space only, while setting VACHT as the primary target, which sampled 31 million stereoisomers and generated 8.8 million compounds. We also ran 100 cycles (2 million rollouts) on the three-component space for the purpose of backpropagation comparison only. Relevant parameters are included in **Table S8-S9**. After the tree search had completed, we collected the compounds in the product layer. For backpropagation evaluation, we only performed a charge filter by removing compounds with formal charge greater than +1.1. For downstream pipeline of hit picking, we performed more elaborate filtering. In the single target setting, we filtered out compounds with undesirable charges and grouped the compounds by SMILES and selected the molecule with the most favorable adjusted docking score for each SMILES to maintain the diversity of the compounds for further hit picking.

Post-docking processing

For σ_2 , the top-ranking 300,000 compounds from the 56-billion-compound CombiDOCK screen were retained for post-docking filtering. To remove molecules similar to known ligands, the ECFP4-based Tc value was computed against 2,232 $\sigma_{1/2}$ ligands in ChEMBL (<https://www.ebi.ac.uk/chembl/>) and 574 σ_2 ligands from S2RSLDB (<https://www.researchdsf.unict.it/S2RSLDB>). Compounds with an Tc ≥ 0.35 were removed, leaving 266,584 molecules. To prioritize poses that capture essential interaction with relevant residues, the remaining molecules were filtered using LUNA-based interaction calculations.¹³⁹ Molecules that capture key interactions with D29 were retained. An additional constraint required fewer than 3 unsatisfied hydrogen bond acceptors and no unsatisfied hydrogen bond donors, narrowing the set to 91,574 compounds. To ensure a diverse set of compounds to select from, the remaining compounds were then clustered by their LUNA interaction fingerprints using the Butina algorithm (Tc ≥ 0.68), and cluster representatives were visually inspected. A total of 137 representatives were selected, and 122 were synthesized and experimentally tested.

The top 300,000 ranked compounds from the MINT-Dock screening were similarly retained for downstream strain filtering (total torsion strain energy of less than 8 units, maximum strain energy per torsion angle of less than 3 units, leading to 229,910 compounds), novelty filtering (Tanimoto similarity with known actives < 0.35 , leading to 178,170 compounds), and interaction filtering with LUNA¹³⁹ on compounds that form a salt-bridge/hydrogen-bond with D29, leading to 68,497 compounds. These filtered compounds were then clustered by their LUNA interaction fingerprints using the Butina algorithm (Tanimoto similarity ≥ 0.68), and cluster representatives were visually inspected. 225 were selected, and 206 were synthesized and tested. Because the same selection criteria were applied, 27 of the 206 compounds tested in the MINT-Dock campaign were also tested in the CombiDOCK campaign: Z9322686774, Z9322686448, Z9322686929, Z9322686891, Z9322686649, Z9322686503, Z9322686489, Z9322687006, Z9322686784, Z9322687611, Z9322686694, Z9322687001, Z9322687500, Z9322686991, Z9322687037, Z9322686403, Z9322687159, Z9322686415, Z9322686618, Z9322686480, Z9322687461, Z9322686557, Z9322687562, Z9322686657, Z9322686646, Z9322687053, and Z9322686947.

For both VMAT2 and VACHT, the top-ranking compounds from each brute-force screen were retained for post-docking filtering (top 100K for 3.6M; top 1M for 1.9B and CombiDOCK 103B). Compounds with an ECFP4 $T_c \geq 0.35$ to known ligands (18 ligands for VMAT2; 38 ligands for VACHT) were removed, leaving 97,740, 978,464, and 993,861 molecules for VMAT2, and 97,693, 958,021, and 983,477 molecules for VACHT, respectively. The remaining molecules were filtered using LUNA-based interaction calculations, retaining only compounds that preserve target-specific bond interaction: for VMAT2, a key hydrogen bond with E312 and an additional hydrogen bond with R189, and a π -stacking interaction with F135; for VACHT, a key hydrogen bond with D398, and at least one additional hydrogen bond with S225, Y432, N302, or N47. An additional constraint required fewer than 3 unsatisfied hydrogen bond acceptors and no unsatisfied hydrogen bond donors, narrowing the sets to 282, 6,127 and 14,701 for VMAT2, and 2,397, 9,356 and 87,428 for VACHT. The filtered compounds were then clustered by their LUNA interaction fingerprints using the Butina algorithm ($T_c \geq 0.75$), and cluster representatives were then visually inspected. Poses were selected for good shape complementarity to the binding site, consistency with established interaction patterns for known binders and excluded if they displayed structural artifacts, such as improper torsional geometries. For VMAT2, preference was given to poses that positioned a protonatable amine toward the conserved acidic residue E312, and maintained compact aromatic scaffolds, and formed well-defined π stacking interactions with F135, while avoiding strained conformations and improper placement of key functional groups. For VACHT, preference was given to poses featuring ring systems over flexible chains, as well as those exhibiting a bidentate binding mode with the binding site. After visual inspection, 78, 122, and 128 compounds were selected for VMAT2 and 83, 130, and 153 for VACHT. For VMAT2, 65, 111, 109 compounds were synthesized and tested experimentally. For VACHT, 64, 108, and 114 were synthesized and tested.

For VMAT2 MINT-Dock screening, we selected the top 1 million ranked compounds for downstream novelty filtering (Tanimoto similarity with 18 known actives < 0.35 , leading to 959,047 compounds) and interaction filtering with LUNA on compounds that form a salt-bridge/hydrogen-bond with E312 and R189 and those that form hydrophobic interaction with F135, leading to 9,028 compounds. These filtered compounds were then clustered by their LUNA interaction fingerprints

using the Butina algorithm (Tanimoto similarity ≥ 0.75), and cluster representatives were visually inspected. Criteria for prioritizing compounds were followed as previously. 124 compounds were selected, and 108 compounds were synthesized and tested. Because the same selection criteria were applied, 17 of the 108 compounds tested in the MINT-Dock campaign were also tested in the CombiDOCK campaign: Z1542734894, Z2481800190, Z8353197938, Z3360769253, Z9329608260, Z9327964514, Z5313551878, Z2839865511, Z3503238340, Z3367013938, Z1839981454, Z4275682064, Z1438338004, Z9332647693, Z9329608325, Z3294417719, Z1272055642.

For selectivity MINT-Dock screening, we filtered out compounds with undesirable charges and grouped the compounds by SMILES and selected the molecule with the largest docking score difference between VACHT and σ_2 for each SMILES before applying another score filter of VACHT docking score < -45 kcal/mol and σ_2 docking score > -40 kcal/mol, leading to 121,628 molecules. For each of these compounds, for pose against each target, we applied strain filtering (111,686 VACHT poses, 110,646 σ_2 poses) and then novelty filtering (111,516 VACHT poses, 107,673 σ_2 poses) as described for σ_2 screening. For interaction filtering using LUNA, we filtered on compounds that form a salt-bridge/hydrogen-bond with the D398 residue of VACHT and do not form an interaction with the D29 residue of σ_2 , leading to 40,449 compounds in total. These filtered compounds were then clustered by their LUNA interaction fingerprints from their docked poses against VACHT using the Butina algorithm (Tanimoto similarity ≥ 0.75), and the VACHT docked poses of cluster representatives were visually inspected. Criteria for prioritizing compounds were followed as previously. 129 compounds were selected, and 114 compounds were synthesized and tested.

For poly-pharmacology MINT-Dock screening, we filtered out compounds with undesirable charges and grouped the compounds by SMILES and selected the molecule with the most favorable sum of docking score for VACHT and docking score for σ_2 for each SMILES before applying another score filter of VACHT docking score < -45 kcal/mol and σ_2 docking score < -45 kcal/mol. For each of these compounds, for pose against each target, we applied strain filtering (111,686 VACHT poses, 110,646 σ_2 poses) and then novelty filtering (111,516 VACHT poses, 107,673 σ_2 poses) as described for σ_2 screening, leading to 260,576 compounds. For interaction filtering using LUNA, we filtered on compounds that both form a salt-bridge/hydrogen-bond with the D398 residue of VACHT and form a salt-bridge/hydrogen-bond with the D29 residue of σ_2 , leading to 80,555 compounds in total. These filtered compounds were then clustered by their LUNA interaction fingerprints from their docked poses against VACHT using the Butina algorithm (Tanimoto similarity ≥ 0.75), and cluster representatives were visually inspected. Compounds with poses satisfying criteria for both the VACHT pocket and the σ_2 pocket were prioritized. 58 compounds were selected, and 53 compounds were synthesized and tested.

Benchmarking CombiDOCK and MINT-Dock against the fragment-enrichment-based approach

To evaluate recall performance of CombiDOCK to V-SYNTHES at large-scale, we curated a benchmark library of ten million molecules generated from 5,000 randomly selected Syn1s and 2,000 Syn2s within a single amide-forming reaction type. Using the top 5000 scoring compounds

(top 0.05% hits) from full-molecule docking as a reference, we assessed the recall performance of CombiDOCK and V-SYNTHES¹⁰ across top ranking thresholds of 0.1%, 0.2%, 0.5%, 1%, 2% and 5% of the ten-million library for the σ_2 receptor, VMAT2 and VACHT. As shown in **Fig. S15A**, CombiDOCK reliably reproduced the top-ranking molecules from full-molecule docking, with recall values for the top 5,000 compounds reaching 45% to 74% at the top 0.2% cutoff of its ranked lists. In contrast, fragment-enrichment approaches, which operate by docking individual synthons to identify promising anchors, enumerating only products containing those anchors, and then docking the reduced product set—achieved recalls of only 4% to 9% using the same docking engine and grids (**Fig. S15A**). Additionally, docking score correlations with full molecule docking remained high across all three targets (σ_2 : Pearson R = 0.89, Spearman R = 0.88; VMAT2: Pearson/Spearman R = 0.67; VACHT: Pearson R = 0.69, Spearman R = 0.70), underscoring strong rank-order performance.

For MINT-Dock, we compared it to the fragment-enrichment protocol above under matched-scale outputs for both σ_2 and VMAT2. For a fair comparison of MINT-Dock with the fragment docking idea of enumerating all products available from top scoring synthons, we re-implemented the V-SYNTHES method as described below. After obtaining the synthon DOCK6 scores, we implemented an automatic pose filter that excluded synthon poses that were likely to clash with the protein. The clash was heuristically determined by extending each synthon pose from the centroid of its cap functional group by a 5 Å distance and checking if the extended coordinate was within a 2.8 Å distance with any heavy atom of the target protein. This 2.8 Å clash distance cutoff derived from the clash criterion established for all-atom contact analysis, where the van der Waals shells of two non-bonded heavy overlapped more than 0.4 Å,¹⁴⁰ was selected as a minimal filter that only excludes synthons when their extension results in a severe steric collision with the target protein. We then selected the top scoring synthons. For both σ_2 and VMAT2, the number of synthons we chose were selected so that the total number of stereoisomers sampled and the total number of compounds generated from fragment docking and MINT-Dock, where compounds from 10 million rollouts on the two-component space and compounds from 2 million rollouts on the three-component space were combined, respectively were similar. Namely, for σ_2 , we pooled 8.3 million two-component compounds (30 million stereoisomers) from 400 cycles (10 million rollout), and 2.0 million three-component compounds (30 million stereoisomers) from 200 cycles (2 million rollout) generated from MINT-Dock, and we pooled 7.6 million two-component compounds (30 million stereoisomers), and 2.6 million three-component compounds (31 million stereoisomers) generated from fragment docking. For VMAT2, we pooled 9.0 million two-component compounds (30 million stereoisomers) from the first 400 cycles (10 million rollout), and 2.0 million three-component compounds (14 million stereoisomers) generated from MINT-Dock, and we pooled 9.1 million two-component compounds (29 million stereoisomers), and 2.9 million three-component compounds (14 million stereoisomers) generated from fragment docking. Since here the purpose was to compare the two methods to evaluate their performances on both dock scores and interactions with the designated target, we performed the same pipeline on the set of molecules generated by fragment docking as we did on the set of molecules generated by MINT-Dock. For each target, we filtered out compounds with undesirable charges and grouped the compounds by SMILES and selected the molecule with the most favorable adjusted docking score for each SMILES. We then selected the top 300,000 ranked compounds, passed them through

strain and novelty filtering, and performed interaction filtering customized for each target. The adjusted enrichment ratio was calculated by the formula presented in comparison with random docking.

For σ_2 (10.3 million molecules from MINT-Dock and 10.2 million from fragment-enrichment docking), fragment-enrichment produced slightly better raw score distributions directly from docking (**Fig. S15B**). However, when we applied interaction-based triage designed to retain molecules consistent with key interactions critical for σ_2 binding, MINT-Dock became increasingly favored. In particular, after filtering for a salt bridge with D29, MINT-Dock outperformed fragment-enrichment at most score thresholds, and the advantage widened further when we imposed a stricter criterion requiring an additional hydrogen-bond interaction (**Fig. S15B**). Composition analysis indicated that fragment-enrichment disproportionately selected high-scoring three-component products, which were less likely to satisfy these interaction filters than two-component products, explaining the reversal after interaction-based triage (**Fig. S16**). For VMAT2 (11.0 million molecules from MINT-Dock and 11.9 million from fragment-enrichment docking), MINT-Dock produced more well-scoring molecules than fragment-enrichment docking across score thresholds, both directly from docking and after applying filters tied to key VMAT2 interactions, including E312 and R189 (**Fig. S15C**).

Given CombiDOCK's higher recall and closer agreement with full docking in the retrospective benchmarking analysis, we next evaluated how a fragment-enrichment-based virtual screening method performs in recovering true actives from our prospective CombiDOCK screens against σ_2 , VMAT2, and VACHT. We used the 103B Enamine REAL library and performed a similar V-SYNTHES-style full molecule enumeration and assessed what percentages of tested actives were included when top 0.02%, 0.05%, 0.1%, 0.5%, 1%, 2%, 5%, and 10% of the top ranked synthons were enumerated.

Fig. S15D summarizes recall at progressively higher selection thresholds for enumerated synthon libraries, where recall values represent the proportion of true actives present in the final library. The fragment-enrichment-based method performed similarly across the three targets, with poor recovery of actives at early stages of enrichment (top $\leq 0.1\%$), with a maximum recall rate of 8% against VACHT if the top 0.1% are enumerated. Recall improved monotonically with broader selections, surpassing 90% by the top 5%. Notably, for VACHT the most potent CombiDOCK hit (CSMS00228273820; IC₅₀ = 12 nM) was not recovered until enumeration of the top 0.5% of synthons, and similarly, for σ_2 and VMAT2 the most potent actives only appeared when the top 5-10% of synthons were selected, requiring enumeration and full docking of more than 500 million molecules (**Table S5**).

With the fragment approach, achieving complete recall of all 64 VACHT actives required enumerating and generating 3D conformers for > 1 billion full molecules. End-to-end, fragment-docking, full-molecule enumeration, followed by full-molecule docking, this takes 179 cluster-days on 2,000 CPUs. In contrast CombiDOCK docks synthons and assembles them into full molecules in a single pass, completing in 26 cluster-days. MINT-Dock achieves even greater efficiency by concentrating docking on the most productive regions of chemical space, typically completing a campaign in 3 to 4 days on approximately 100 CPU cores while generating molecules that better

satisfy target-specific interaction criteria than fragment-enrichment at matched scale. Together, these comparisons establish that both CombiDOCK and MINT-Dock overcome the core limitation of fragment-enrichment approaches.

DrugCLIP comparison

The recently published DrugCLIP is able to score over trillions of ligand-protein interactions per cluster day using 8 GPUs, which has enabled genome-wide docking, discovering potent binders prospectively for three biological targets.⁶ DrugCLIP reports that screening 500 million compounds against 10,000 proteins takes 1 cluster-day on 8 GPUs. However, under DrugCLIP's stated computational complexity $O(M+N)$ (with M targets and N compounds), scaling from 0.5 billion to 100 billion compounds implies 200 cluster-days on the same 8-GPU node, largely independent of whether M is 10,000 proteins or a single protein (because runtime is dominated by N). In addition, DrugCLIP's ligand encoder (Uni-Mol¹⁴¹) requires precomputed 3D conformers generated using RDKit, and the paper describes a data augmentation strategy that can generate up to 10 conformations per molecule. Even assuming a conservative one conformer per molecule for library embedding, conformer generation becomes a major overhead at the 100-billion scale. For example, with 2,000 CPU cores and 0.2 s per conformer per CPU-core, generating one conformer per molecule for 100B molecules requires 116 cluster-days. Taken together, even under optimistic assumptions, the end-to-end runtime for the DrugCLIP retrieval step at 100 billion molecules on a single target is at least 316 cluster-days. In contrast, CombiDOCK requires only 1 cluster-day to prepare the synthon library and 30 cluster-days to screen 100 billion molecules against a single target, implying roughly a 10-fold faster end-to-end workflow at this scale. Moreover, DrugCLIP is fundamentally a retrieval approach: to make structure-based predictions, it must be paired with additional docking engines, which contributes to end-to-end wall time beyond the retrieval stage. On our targets, DrugCLIP predictions do not even recover the experimental ligand binding site. For σ_2 , DrugCLIP yields no predictions in [their public database](#), which is surprising given σ_2 is considered ligandable. For both VMAT2 and VACHT, visual inspection of the reported DrugCLIP pockets/hits indicates all predicted binding sites that are spatially distant from the canonical ligand site (**Fig. S11C-D**). We therefore re-docked the top DrugCLIP hits using the same docking engine used in this study and found they score substantially worse than CombiDOCK's top-ranked compounds on both VMAT2 and VACHT (**Fig. S11C-D**).

Analoging in the make-on-demand library

The most potent compound identified in our 103 billion CombiDOCK screen against VACHT was CSMS00228273820 with an $IC_{50} = 12$ nM. The most potent compound identified in our MINT-Dock screen for selectivity was Z9424360192 and Z2406174633 with an $IC_{50} = 86$ nM and 100 nM respectively. To optimize the potency of the three compounds we curated a set of close structural analogues from Enamine REAL Space make-on-demand library for testing. Using database search tool swp.docking.org,¹⁴² we queried 68 compounds with a maximum graph edit distance of 12 relative to the structure of CSMS00228273820, 1176 compounds with a maximum graph edit distance of 6 relative to the structure of Z9424360192 and 1128 compounds with a maximum graph edit distance of 7 relative to the structure of Z2406174633. Stereoisomers and

protomers were enumerated for all compounds followed by 3D conformer generation. Conformers for all compounds were stored in DB2 format and were docked to both our VACHT and σ_2 models using DOCK6 HDB. Here we placed the primary focus on structural similarity of analogs with our parent compounds, with σ_2 selectivity as a secondary filter. For CSMS00228273820, we selected 21 analogues for synthesis, all of which were successfully synthesized and experimentally tested. For Z9424360192, we selected 22 analogs for synthesis, 20 were successfully synthesized and experimentally tested. For Z2406174633, we selected 32 analogs for synthesis, 29 were successfully synthesized and experimentally tested. Of these 29 analogs, Z2406174639, Z2406174433, and Z2405924153 were also analogs of CSMS00228273820.

Expression and purification of VACHT

The coding sequence corresponding to human VACHT (residues 34 to 524) was subcloned into the pBMCL1 vector.¹⁴³ The DNA sequence encoding MBP and the helical linker with the sequence (AEEERK)₂ were incorporated at the N terminus of VACHT. Meanwhile, the flexible linker (GGGGS)₂ was fused to the C-terminus of VACHT, together with the coding sequence for DARPinoff₇¹⁴⁴ (TransGen Biotech). This fully assembled construct was referred to as the VACHT^{EM} hereafter.

VACHT^{EM} was expressed in HEK293S GnTI⁻ cells utilizing the BacMam system.¹⁴⁵ Baculoviruses were generated through transfection of Sf9 cells. Following three successive rounds of amplification, the viruses were used for cell transduction. HEK293S cells, maintained at a density of 3×10^6 cells/ml, were subjected to infection via the addition of 10% third-passage baculoviruses, and subsequently cultured at 37°C for 8–12 h. Thereafter, 10 mM sodium butyrate was introduced to trigger protein expression, after which the cultures were further incubated at 30°C for an additional 48 h. The cells were harvested via centrifugation at 4,000 rpm for 20 min and then stored at -80°C.

To purify VACHT^{EM}, cell pellets were resuspended in lysis buffer (50 mM HEPES pH 7.25, 300 mM NaCl, 15% glycerol) supplemented with 2 μ g/ml DNase I and protease inhibitor cocktail (APEX-BIO). Cell lysis was then performed at 4°C for 3 h using 2% Lauryl maltose neopentyl glycol (LMNG), 0.2% Cholesteryl hemisuccinate (CHS), and 0.67% Glyco-dendrimer (GDN). Following centrifugation at 18,000 rpm for 40 min, the supernatant was subjected to incubation with anti-GFP nanobody affinity resin at 4°C for 2 h. The resin was subsequently washed using 40 column volumes of Buffer A (25 mM HEPES pH 7.25, 150 mM NaCl, 0.003% LMNG, 0.001% GDN, 0.0003% CHS). To remove the GFP tag and release VACHT^{EM} from the resin, GST-tagged PreScission Protease was utilized. The protease was thereafter removed via adsorption to Glutathione Beads (Smart-Lifesciences). The protein underwent further purification via size-exclusion chromatography (SEC), using a Superose 6 Increase 10/300 GL column (GE Healthcare) in Buffer A. Peak fraction containing VACHT^{EM} was concentrated to a final concentration of ~8 mg/ml using a 50 kDa cut-off centrifugal filter (Millipore). The protein sample was stored at -80°C until subsequent use.

Cryo-EM sample preparation and data collection

The protein was pre-incubated with 600 μM CSM00228273820 or Z9424360192 at 4°C for 20 min before initiating cryo-EM sample preparation. A 3 μL aliquot of the protein sample was applied onto glow-discharged holey carbon grids (Quantifoil R0.6/1.0, Au300 mesh). Using a Vitrobot Mark IV (FEI), the grids were flash-frozen in liquid ethane cooled via liquid nitrogen under conditions of 8°C and 100% humidity. For the blotting step, the operational parameters were set as follows: a blot time of 3 s and a wait time of 10 s.

The grids were screened utilizing a 200 kV Talos Arctica microscope (FEI), which was outfitted with a Gatan K2 Summit detector. Acquisition of raw movie stacks was carried out with a 300 kV Titan Krios microscope (FEI) paired with a Gatan K3 camera. The data were recorded at a physical pixel dimension of 0.83 Å per pixel and a nominal defocus interval ranging from 1 to 2 μm , with each individual movie comprising 40 frames. The total exposure dose was around 60 $\text{e}^-/\text{Å}^2$. Ultimately, 9,261 and 6,847 micrographs were acquired in total for CSM00228273820-bound VACHT and Z9424360192-bound VACHT respectively. Comprehensive details of all data collection parameters are compiled in **Table S4**.

Cryo-EM image processing

All data processing procedures were carried out solely within cryoSPARC.¹⁴⁶ The image stacks were gain-normalized and corrected for beam-induced motion using Patch Motion Correction. Estimation of CTF parameters was accomplished through the Patch CTF Estimation tool. Micrographs that were unsuitable for subsequent analysis were excluded via manual inspection. Blob picking was utilized for initial particle picking and the generation of 2D templates. Following three cycles of 2D classification, particles of high quality were used for ab initio reconstruction and heterogeneous refinement processes. Low-pass filtered maps obtained from this stage guided further heterogeneous refinement, which was subsequently followed by non-uniform refinement¹⁴⁷ of particles derived from the optimal classes. These refined particles served as reference templates for template-based picking or as training datasets for the Topaz particle-picking pipeline.¹⁴⁸ Subsequent iterations of ab initio reconstruction, heterogeneous refinement, and additional heterogeneous refinement steps, guided by low-pass filtered maps from the prior blob picking step, were repeated to merge the optimal particle classes. Following the removal of duplicate particles, the resulting particle subsets underwent refinement via both non-uniform refinement and local refinement. Subsequently, the refined map and particles were subjected to seed-facilitated classification and refinement,¹⁴⁹ with this process conducted against the full dataset of initially picked particles. A final round of local refinement, focused on the transporter, yielded a 3.6 Å-resolution map for the CSM00228273820-bound state and a 3.7 Å-resolution map for the Z9424360192-bound state (gold-standard FSC = 0.143).

Model building and refinement

The vesamicol-bound VACHT structure (PDB: 8ZMR)¹⁰¹ was first docked into the cryo-EM density map using ChimeraX.¹⁵⁰ Subsequently, real-space refinement was executed via PHENIX,¹⁵¹ while manual adjustments were made within Coot¹⁵². Iterative cycles of automatic and manual refinement were conducted until the final structural model was achieved. Estimation of the local resolution for the cryo-EM map was carried out using cryoSPARC.¹⁴⁶ The geometric integrity of

the structural model was validated through MolProbity.¹⁴⁰ All structure-associated figures were generated utilizing ChimeraX.

σ_2 radioligand binding assay

Membranes for radioligand binding assays were prepared from Expi293 cells (Thermo Fisher Scientific) transfected with wild-type σ_2 receptor using FectoPRO (Polyplus-transfection) according to the manufacturer's instructions. Cells were harvested 72 hours post-transfection and pelleted by centrifugation. Cell pellets were resuspended in 20 mM HEPES (pH 7.5), 2 mM magnesium chloride, and 1:100,000 (vol:vol) benzonase nuclease (Sigma Aldrich), supplemented with cOmplete Mini EDTA-free Protease Inhibitor Mixture Tablets (Roche). Following Dounce homogenization, cells were centrifuged at 50,000 × g for 20 min at 4°C. The resulting membrane pellet was resuspended in 50 mM Tris (pH 8.0), subjected to final Dounce homogenization, aliquoted, flash frozen, and stored at −80°C until use.

For radioligand binding experiments membranes were thawed, homogenized, and incubated with 5 nM [³H]-DTG (PerkinElmer) in 100 μ L binding reactions containing 50 mM Tris (pH 8.0), 0.1% (w/v) bovine serum albumin, and 50 nM PD-144418 to block σ_1 receptor binding sites. For competition assays, competing ligands were added at 1 μ M for single-point measurements. Reactions were incubated at 37°C for 2 hours with gentle agitation. Reactions were terminated by rapid filtration through glass fiber filters pre-treated with 0.3% polyethylenimine using a Brandel cell harvester. Filters were washed with ice-cold water, then individually soaked in 5 mL Cytoscent scintillation fluid (MP Biomedicals) overnight. Radioactivity was quantified using a Beckman Coulter LS6500 scintillation counter. All reactions were performed in triplicate using 96-well format.

VMAT2 radioligand uptake assay

The uptake assays were performed as previously described.⁹⁸ Human embryonic kidney (HEK293) cells (ATCC, CRL-3216) were transiently transfected with a VMAT2-encoding plasmid using Lipofectamine 3000 according to the manufacturer's protocol. Cells were cultured at 37 °C for 16–18 h, supplemented with 10 mM sodium butyrate, and further incubated at 30 °C for an additional 24 h.

Assays were performed 48 h after transfection. The medium was aspirated, and cells were washed with ice-cold assay buffer (110 mM potassium gluconate, 5 mM glucose, 5 mM MgCl₂, 1 mM ascorbic acid, 30 μ M pargyline and 20 mM HEPES pH 7.4), then incubated with 10 μ M digitonin in assay buffer for 5 min at room temperature and washed again with ice-cold assay buffer without digitonin. Permeabilized cells were incubated with an assay buffer containing 5 mM ATP with test compounds at a final concentration of 10 μ M. After 10 min incubation at room temperature, 50 nM [³H]-dopamine (ViTrax Radiochemicals) was added, cells were subsequently incubated at room temperature for 20 min. The substrate-containing buffer was aspirated, cells were washed twice with ice-cold assay buffer and then solubilized in P2 lysis buffer (Qiagen). Scintillation liquid (Ultima Gold) was added to cell lysates to measure radioactivity using a Microbeta counter (PerkinElmer).

VACHT radioligand binding assay

Radioligand binding experiments were performed utilizing crude cellular membranes harboring human VACHT. HEK293F cells that had been transiently transfected with full-length VACHT plasmid were cultured at 37°C for 16–18 h, after which 10 mM sodium butyrate was added and the cells were further incubated at 30°C for additional 24 h. Following cell collection, the cells were washed once with cold PBS and then resuspended in ice-cold SH buffer (10 mM HEPES pH 7.4, 320 mM sucrose, protease inhibitor cocktail) at a density of approximately 4×10^6 cells/mL. The resuspended cells were lysed via sonication, then centrifuged at 4,000 g for 5 min to eliminate unlysed cells and cellular debris. The resulting supernatant was then flash-frozen in liquid nitrogen and stored at –80°C for subsequent use.

In the [3H]-vesamicol competition binding assay, 50 μ L aliquots of membrane suspensions were subjected to a 60-min incubation at 25°C with 150 μ L of binding buffer (20 mM HEPES pH 7.4, 110 mM potassium tartrate, 1 mM ascorbic acid) containing 3 nM (–)-[3H]-vesamicol (30–60 Ci/mmol, American Radiolabeled Chemicals, Inc.) and 7.5 μ M of the test compounds. Reactions were terminated via rapid vacuum filtration through GF/C glass fiber filters (Whatman) pretreated with 0.3% polyethyleneimine, with subsequent washing using 2 ml of ice-cold binding buffer. The radioactivity retained on the filters was quantified by scintillation counting in 2.5 mL of scintillation solution (PerkinElmer).

Dose response curve

For σ_2 top compounds measurement, the dose-response assays were performed in the same manner as the competition binding assays, except that test compounds were added at a range of concentrations to generate full displacement curves. Data were analyzed using GraphPad Prism 10 (<http://www.graphpad.com>). K_i values were calculated using the Cheng–Prusoff equation as implemented in the software.

For VMAT2 top compounds measurement, the dose-response assays were performed similarly to the [3H]-dopamine uptake assay, except that the permeabilized cells were incubated with assay buffer containing 5 mM ATP with a series of specified concentrations of the test compounds. IC_{50} values were determined by fitting three-parameter dose–response curves using GraphPad Prism 10.

For VACHT top compounds measurement, the dose-response assays were conducted in a similar manner to the competition binding assay, with the exception that the membrane suspension was incubated with a series of specified concentrations of the test compounds. The IC_{50} values were determined by fitting $logIC_{50}$ curves using GraphPad Prism 10.

Microscale Thermophoresis (MST) analysis

The binding affinity between VACHT (and its mutants) and CSMS00228273820 was measured by MST experiments using a NanoTemper Monolith NT.115 instrument (NanoTemper Technologies). VACHT and mutants were first fluorescently labeled using the Red-NHS 2nd

Generation kit (NanoTemper Technologies) in the assay buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 0.02% DDM and 0.002% CHS). For the binding assay, 50 nM of the labeled VACHT was then incubated with CSMS00228273820 at 16 serially diluted concentrations, ranging from 0.476 nM to 15.6 μ M, prepared in the same assay buffer. After a 20-min incubation at 16°C, the mixture was then loaded into Standard Monolith Capillaries (NanoTemper Technologies). MST measurements were conducted using an excitation power of 20% and an MST power of 40%. Capillaries exhibiting aggregation or adsorption were excluded. The K_d value was calculated using the MO.Affinity Analysis software (NanoTemper Technologies).

Make-on-demand synthesis

All the make-on-demand molecules were derived from Enamine REAL (<https://enamine.net/compound-collections/real-compounds>). See supplementary materials for the synthesis procedures and characterization of compounds.

Bemis-Murcko scaffold analysis

Molecular scaffolds were generated using RDKit,¹²⁸ following the Bemis–Murcko framework,¹¹⁶ to enable structural comparisons among compounds from different library screens.

MINT-Dock comparison with large-scale docking

The large-scale docking campaigns for 3 targets were described previously. We added another σ_2 docking campaign of 1.73 billion stereogenic molecules, which is a strict subset of the 1.85 billion molecules in prior VMAT2 and VACHT screening campaigns, for both an additional comparison with MINT-Dock and comparison with multi-target MINT-Dock. The less 0.12 billion molecules were not procured due to a reorganization of the ZINC22 database. For rank comparison, we coalesced the stereoisomers and assigned the best DOCK score for each unique ZINC ID from the large-scale docking campaign. For single-target comparison, we used the formula $\text{rank} = \frac{\sum_{i \in \text{Docking Campaign}} I(s_i < s_0)}{\sum_{i \in \text{Docking Campaign}} 1}$ that is the fraction of docked molecules for each docking campaign with scores below a threshold s_0 to represent its efficiency. The higher the rank is, the more efficient the campaign is. For VMAT2 comparison, we applied a post-docking strain filter of total torsion strain energy of less than 8 units, maximum strain energy per torsion angle of less than 1.5 units to molecule poses from the 1.85 billion large-scale docking campaign in order to match with the MINT-Dock docking setup. The 660-fold enrichment for σ_2 was determined from using the score threshold of -56.5 kcal/mol, and the 130-fold enrichment for VMAT2 was determined from using the score threshold of -53.5 kcal/mol, same as in the “**MINT-Dock Backpropagation Comparison**” section. For rank comparison for selectivity screens, we used the formula $\text{rank} = \frac{\sum_{i \in \text{Filtered Mol Set}} I(s_i < s_0)}{\sum_{i \in \text{Docking Campaign}} 1}$, where “Filtered Mol Set” was defined to be the set of molecules with VACHT docking score below -45 kcal/mol and σ_2 docking score above -40 kcal/mol. For rank comparison for poly-pharmacology screens, we used the formula $\text{rank} = \frac{\sum_{i \in \text{Filtered Mol Set}} I(s_i < s_0)}{\sum_{i \in \text{Docking Campaign}} 1}$, where “Filtered Mol Set” was defined to be the set of molecules with VACHT docking score below -45 kcal/mol and σ_2 docking score below -45 kcal/mol.

Technical details for CombiDOCK

CombiDOCK was implemented in the latest release of DOCK6 (version 6.14). It was built upon a hierarchical database (HDB) search routine, which traverses precomputed ligand conformations stored in a Dock Binary version 2 (DB2) format, in a similar way to that in DOCK3.^{50,127} Our method is based on HDB docking framework re-implemented in DOCK6⁸⁴ but extends them by enabling hierarchical tree sampling of two molecular fragments in a sequential and combinatorial manner. Below are the technical details:

1. DB2 file format in hierarchical organization

The DB2 file format stores ligand conformations in a highly compressed and hierarchical structure. In this format, each molecule is associated with one or more hierarchies, each built around a rigid segment, which serves as a reference frame for all conformational variations. For example, a molecule containing multiple rings will have each ring treated as an independent rigid segment, and a separate hierarchy will be generated for each. Thus, a molecule with three rings will have three corresponding hierarchies.

Each hierarchy is organized as a tree structure, where the root node represents the rigid segment, and subsequent nodes represent molecular segments in specific conformations. A complete branch of the tree corresponds to a fully grown molecular conformation. Since the rigid root segment is shared among all branches, it provides a fixed reference for the relative positions of different branches.

The standard DB2 file¹²⁷ is composed of several sections:

- M (Header): General molecule information
- A (Atoms): Atom definitions
- X (Coordinates): Cartesian coordinates of atoms
- C (Conformers): Molecular segment information
- S (Sets): Branch or conformer information

This hierarchical organization enables efficient storage and rapid conformational search, making the DB2 format well-suited for ultra-large library docking. To support combinatorial docking of two synthons, CombiDOCK extends the DB2 format with an additional overlay section:

- O (Overlay): Overlay segment information

The overlay section contains the overlay segment information, which is derived from its capping group. This information is used to align and fuse two separate fragments into a complete molecule during docking. By including an overlay section within the DB2 structure, CombiDOCK maintains the efficiency of hierarchical sampling and full molecule reconstruction.

Moreover, because each synthon can serve as either the first (Syn1) or the second (Syn2) fragment in the docking process, we prepared two separate DB2 files for each: In the Syn1 form, the number of ring systems determines the number of hierarchies. Each ring is treated as a rigid segment and used as the root of a hierarchy. In the Syn2 form, only a single hierarchy is

generated, with the capping group defined as the rigid segment. This dual representation ensures that each synthon can be appropriately used for both stages of combinatorial docking.

2. DB2 building workflow

The DB2 building workflow converts 2D molecular representations (full molecules or synthons) into 3D conformer ensembles stored in the hierarchical DB2 format, enabling efficient sampling during docking. The process involves six key stages: protonation and tautomerization state enumeration, stereoisomer enumeration, 3D conformer generation, partial charge and desolvation calculation, conformational expansion, and hierarchical data structuring.

- **Protonation and tautomerization state enumeration**
Each input molecule (represented using SMILES string) is first processed using the Jchem suite (v.22.22.0, <https://chemaxon.com/>). The cxcalc and molconvert utilities are used to enumerate protonation and tautomerization states at physiological pH (7.4). Each unique protomer or tautomer is treated as an independent molecular entry for downstream processing.
- **Stereoisomer enumeration**
For each protomer/tautomer, RDKit (v2023.09.1)¹²⁸ is used to enumerate all possible stereoisomers based on undefined chiral centers. Each stereoisomer is retained as an independent molecular variant for downstream 3D conformer generation.
- **3D conformation generation**
Each stereoisomer is converted into a 3D geometry using Corina Classic (v.4.4.0026, <https://mn-am.com/products/corina/>), which generates a single low-energy conformation while preserving stereochemistry. This initial geometry serves as the input for subsequent partial charge and solvation calculations.
- **Partial charge and desolvation calculations**
AMSOL (version 7.1) is used to compute AM1-BCC partial atomic charges^{153,154} and per-atom desolvation energy decompositions using a Generalized Born (GB) solvation model. These charge and solvation descriptors are stored with the molecular data (mol2 format) and propagated into the DB2 structure for electrostatic and ligand solvation terms.
- **Conformational expansion**
Using the Corina-generated 3D geometry as a reference, OMEGA (v.2019.10.2, <https://www.eyesopen.com/omega>)¹²⁹ is employed to generate a hierarchical ensemble of conformers by systematically sampling torsional degrees of freedom. The resulting conformers are filtered by energy and RMSD thresholds (energy window = 12, RMSD cutoff = 0.5 Å) to remove redundant or high-energy conformations, yielding a compact yet representative conformational set.
- **Hierarchy construction and DB2 generation**
The OMEGA-generated conformers are organized hierarchically according to their rigid segments using the mol2db2 utility from the DOCK 3.7 (<https://dock.compbio.ucsf.edu/DOCK3.7/>). Each rigid segment serves as the root of a hierarchy, while flexible torsions define successive branches.
 - For multi-ring full molecules, each ring is treated as a separate rigid root, resulting in multiple hierarchies.

- For synthons, two DB2 representations are generated:
 - Syn1 form – multiple hierarchies, one per ring system.
 - Syn2 form – a single hierarchy rooted at the capping group.

This dual representation allows each synthon to function as either the first (Syn1) or second (Syn2) fragment in the combinatorial docking process.

Hierarchical data are converted into DB2 format using the `build_ligands.py` utility from the DOCK 3.7. Each DB2 file contains five primary sections – M (Header), A (Atoms), X (Coordinates), C (Conformers), and S (Sets) – and, for synthons, an additional O (Overlay) section encoding atom indices for fragment fusion during docking. We modified `build_ligands.py` to incorporate RDKit for reading 3D molecular structures (mol2 files) and automatically detecting atom indices corresponding to the capping group, which enables accurate Overlay information for combinatorial docking.

3. HDB search routine

The HDB search begins by aligning the rigid segment of the ligand to the receptor matching spheres. If the alignment is incompatible (e.g., score is not acceptable due to the steric clashes), the match is discarded, and the next alignment is considered.

For each conformational set (or branch set), segment conformations are evaluated sequentially. Each segment is first oriented based on the rigid segment position and then scored. If a segment is found to be incompatible (e.g., score is not acceptable due to the steric clashes), it is flagged. This prevents redundant evaluation if the same segment is encountered in other branches. Once an incompatibility is detected, the current search along these sub-branches is terminated, and the routine proceeds to the next branch. If all segment conformations in a branch are compatible, the entire branch is retained as a valid conformation. After all the branches are traversed, each valid conformation is minimized through rigid-body minimization to further refine the pose. The final refined poses are ranked by docking scores and the top-scoring pose is retained for virtual screening (VS).

We implemented CombiDOCK HDB search routine by enabling hierarchical search of two molecular fragments in a sequential and combinatorial manner. Same as in the standard HDB search, the rigid segment of the first synthon (Syn1) is oriented to the receptor matching spheres, and segment conformations are evaluated sequentially along each branch. Incompatibilities of the segment conformation terminate the evaluation of that branch, and the routine continues to the next until all branches are traversed. Valid Syn1 conformations are saved.

Next, for each valid Syn1 conformation, its capping group is used as the reference frame to orient the rigid segment of the second synthon (Syn2). The rigid segment of Syn2 is aligned by superposition of its capping group onto that of Syn1 using the Kabsch algorithm. After alignment, the search proceeds using the same segment-by-segment evaluation described above. All valid Syn2 conformations are stored for each Syn1 pose, and the resulting combinations are fused into complete molecules. These full-molecule poses are refined through rigid-body minimization and then ranked by docking score. The top-scoring pose is retained for VS.

4. Fuse Syn1-Syn2 into full molecule

Each full molecule is constructed by fusing a pair of synthons (Syn1 and Syn2) using predefined overlay information extracted from DB2 files. This overlay section specifies the atom indices to be deleted from Syn1 and Syn2, as well as the atom indices from each synthon to be bonded to form the final molecule.

During the construction of the full molecule, atomic partial charges are rescaled using a sign-correcting approach. For example, if the computed total charge of the resulting molecule is +0.2 but the target total charge should be 0, the charge corrector is -0.2 (i.e., $0 - 0.2 = -0.2$). Since this corrector is a negative value, only the negatively charged atoms are uniformly adjusted. Each negative atom receives a per-atom correction equal to the total charge corrector divided by the number of negatively charged atoms. This adjustment slightly increases the magnitude of the negative charges while preserving the overall charge distribution. In parallel, the atomic polar solvation energies are rescaled proportionally based on the ratio of corrected to original partial charges. This ensures consistency in energy values following charge adjustment.

5. Combinatorial docking parameters

The docking parameters used for CombiDOCK are summarized in the table below and parameters specific to the CombiDOCK are highlighted in **Table S6**.

Data availability

Active molecules reported here are available from J.L. or directly from Enamine (enaminestore.com). The structures of VACHT determined in complex with docking hits have been deposited in the Protein Data Bank under accession codes PDB: XXX (CSMS00228273820-bound) and PDB: XXX (Z9424360192-bound), and the corresponding cryo-EM maps are available from the Electron Microscopy Data Bank under accession codes EMD-XXXXX and EMD-XXXXX, respectively. The compounds docked from the ZINC make-on-demand library are freely available at <http://zinc22.docking.org>. Docking scores and SMILES for all screened molecules have been deposited on the Lyu Lab (Rockefeller University) Globus Share Collection: CombiDOCK_docking_data, and will be made publicly available upon publication. Any other data relating to this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

Code availability

DOCK 3.8 code is freely available on [GitHub](#). CombiDOCK is implemented in DOCK 6.13. The code will be freely available on [GitHub](#) upon publication. The MINT-Dock code will be available on [GitHub](#) upon publication.

Acknowledgements

This work was supported by Rockefeller University start-up funds (J.L.), Irma T. Hirschl/Monique Weill-Caulier Trusts (J.L.), and the Stavros Niarchos Foundation (SNF) as part of its grant to the SNF Institute for Global Infectious Disease Research at The Rockefeller University (J.L.). J.L. is supported by the Searle Scholars Program. This research was also funded by the National Natural Science Foundation of China (to Z.Z., NO. 32471252) and the Beijing Natural Science Foundation (to Z.Z., NO. Z240013). The study was also supported in part by the Center for Life Sciences and the Qidong-SLS Innovation Fund (to Z.Z.). We thank the High-Performance Computing Resource Center (HPCRC) at The Rockefeller University (RRID:SCR_025889) for computational resources. We thank Trent Balias for assistance with the DOCK6 source code and helpful discussions; Aakash Dadasam for providing AF3 benchmarking data for σ_2 and SLC transporters; Dr. Jason Banfelder and Dr. Linelle Abueg for testing MINT-Dock code at Rockefeller HPCRC. We thank the Cryo-EM platform at the School of Life Sciences (SLS) of Peking University for helping with cryo-electron microscopy data collection. We thank Dr. Shitang Huang at the National Center for Protein Sciences of Peking University for assistance with the radioactive [^3H]-vesamicol displacement experiments. We thank Dr. Brian Shoichet, Dr. Tarun Kapoor, Dr. Jue Chen, Dr. Seth Darst, Dr. Trent Balias, Dr. Jie Zhou, Dr. Fangyu Liu, and Dr. Hongyan Du for reading this work and helpful suggestions. We thank ChemAxon for a license to JChem, OpenEye Scientific software for a license to OEChem and Omega2, Molecular Networks for a license to Corina, Schrodinger LLC for the use of the prepwizard program in Maestro, Revvity Signals for the use of ChemDraw. Molecular graphics and analyses were performed with UCSF Chimera and UCSF ChimeraX. We polished and revised the manuscript with ChatGPT-5.3 and ChatGPT-5.4. We thank BioRender.com for graphic elements that were incorporated in our Figure 1B and 1F.

Author contributions

J.L. conceived the study; C.Y. developed and implemented the CombiDOCK method in DOCK 6 and built the DOCKable files for the Synthon library; J.Z. and J.L. conceptualized the MINT-Dock algorithm; J.Z. developed the MINT-Dock algorithm; J.L. constructed docking grids for VMAT2 and VACHT, and performed subsequent in-stock and ZINC22 docking; C.Y. and B.L. performed

CombiDOCK docking, chemoinformatics analyses, and ligand selection, with assistance from J.L.; J.Z. ran computational benchmarking for MINT-Dock with assistance from B.L.; J.Z. ran and analyzed MINT-Dock prospective campaigns with assistance from C.Y., B.L., and J.L.; Y.W. and C.Y. organized the large-scale docking data; Y.Z. and Z.Z. performed primary radioligand displacement assays, dose-response curve measurements, and cryo-EM structure determination for VACHT compounds; C.B. and A.A. performed primary radioligand displacement assays and dose-response curve measurements for compounds from the σ_2 campaign and VACHT multi-target campaigns; X.C., S.P., and C.L. performed primary radioligand displacement assays and dose-response curve measurements for VMAT2 compounds; Y.S.M. and D.S.R. directed compound synthesis, purification, and characterization; J.L., Z.Z., C.L., and A.A. supervised the project; J.L., C.Y., J.Z., and B.L. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

References

1. Lyu, J., Wang, S., Balius, T.E., Singh, I., Levit, A., Moroz, Y.S., O'Meara, M.J., Che, T., Algaa, E., Tolmachova, K., et al. (2019). Ultra-large library docking for discovering new chemotypes. *Nature* 566, 224–229. <https://doi.org/10.1038/s41586-019-0917-9>.
2. Irwin, J.J., Tang, K.G., Young, J., Dandarchuluun, C., Wong, B.R., Khurelbaatar, M., Moroz, Y.S., Mayfield, J., and Sayle, R.A. (2020). ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 60, 6065–6073.
3. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* 54, 5.6.1-5.6.37. <https://doi.org/10.1002/cpbi.3>.
4. Lyu, J., Kapolka, N., Gumpfer, R., Alon, A., Wang, L., Jain, M.K., Barros-Álvarez, X., Sakamoto, K., Kim, Y., DiBerto, J., et al. (2024). AlphaFold2 structures guide prospective ligand discovery. *Science* 384, eadn6354. <https://doi.org/10.1126/science.adn6354>.
5. Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T.J., Berenberg, D., Fisk, I., Zanichelli, N., et al. (2024). OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat Methods* 21, 1514–1524. <https://doi.org/10.1038/s41592-024-02272-z>.
6. Jia, Y., Gao, B., Tan, J., Zheng, J., Hong, X., Zhu, W., Tan, H., Xiao, Y., Tan, L., Cai, H., et al. (2026). Deep contrastive learning enables genome-wide virtual screening. *Science* 391, eads9530. <https://doi.org/10.1126/science.ads9530>.
7. Seffernick, J.T., and Fischer, M. (2025). An experimental proxy of water displaceability for ligand discovery. *Nat Methods* 22, 1476–1485. <https://doi.org/10.1038/s41592-025-02724-0>.
8. Kuan, J., Radaeva, M., Avenido, A., Cherkasov, A., and Gentile, F. (2023). Keeping pace with the explosive growth of chemical libraries with structure-based virtual screening. *WIREs Computational Molecular Science* 13, e1678. <https://doi.org/10.1002/wcms.1678>.
9. Sadybekov, A.V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685.
10. Sadybekov, A.A., Sadybekov, A.V., Liu, Y., Iliopoulos-Tsoutsouvas, C., Huang, X.-P., Pickett, J., Houser, B., Patel, N., Tran, N.K., Tong, F., et al. (2022). Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* 601, 452–459. <https://doi.org/10.1038/s41586-021-04220-9>.
11. Petrović, D., Scott, J.S., Bodnarchuk, M.S., Lorthioir, O., Boyd, S., Hughes, G.M., Lane, J., Wu, A., Hargreaves, D., Robinson, J., et al. (2022). Virtual Screening in the Cloud Identifies Potent and Selective ROS1 Kinase Inhibitors. *J. Chem. Inf. Model.* 62, 3832–3843. <https://doi.org/10.1021/acs.jcim.2c00644>.
12. Beroza, P., Crawford, J.J., Ganichkin, O., Gendele, L., Harris, S.F., Klein, R., Miu, A., Steinbacher, S., Klingler, F.-M., and Lemmen, C. (2022). Chemical space docking enables large-scale structure-based virtual screening to discover ROCK1 kinase inhibitors. *Nat. Commun.* 13, 6447.

13. Müller, J., Klein, R., Tarkhanova, O., Gryniukova, A., Borysko, P., Merkl, S., Ruf, M., Neumann, A., Gastreich, M., Moroz, Y.S., et al. (2022). Magnet for the Needle in Haystack: “Crystal Structure First” Fragment Hits Unlock Active Chemical Matter Using Targeted Exploration of Vast Chemical Spaces. *J. Med. Chem.* *65*, 15663–15678. <https://doi.org/10.1021/acs.jmedchem.2c00813>.
14. Lutzens, A., Vo, D.D., Scaletti, E.R., Wiita, E., Almlöf, I., Wallner, O., Davies, J., Košenina, S., Meng, L., Long, M., et al. (2025). Virtual fragment screening for DNA repair inhibitors in vast chemical space. *Nat Commun* *16*, 1741. <https://doi.org/10.1038/s41467-025-56893-9>.
15. Ghazi Vakili, M., Gorgulla, C., Snider, J., Nigam, A., Bezrukov, D., Varoli, D., Aliper, A., Polykovsky, D., Padmanabha Das, K.M., Cox III, H., et al. (2025). Quantum-computing-enhanced algorithm unveils potential KRAS inhibitors. *Nat Biotechnol* *43*, 1954–1959. <https://doi.org/10.1038/s41587-024-02526-3>.
16. Baidya, A.T., Goswami, A.K., Das, B., Darreh-Shori, T., and Kumar, R. (2024). AI-enabled ultra-large virtual screening identifies potential inhibitors of choline acetyltransferase for theranostic purposes. *ACS Chem. Neurosci.* *15*, 4156–4170.
17. Bos, P.H., Houang, E.M., Ranalli, F., Leffler, A.E., Boyles, N.A., Eyrich, V.A., Luria, Y., Katz, D., Tang, H., Abel, R., et al. (2022). AutoDesigner, a De Novo Design Algorithm for Rapidly Exploring Large Chemical Space for Lead Optimization: Application to the Design and Synthesis of d-Amino Acid Oxidase Inhibitors. *J. Chem. Inf. Model.* *62*, 1905–1915. <https://doi.org/10.1021/acs.jcim.2c00072>.
18. Zhou, G., Rusnac, D.-V., Park, H., Canzani, D., Nguyen, H.M., Stewart, L., Bush, M.F., Nguyen, P.T., Wulff, H., Yarov-Yarovoy, V., et al. (2024). An artificial intelligence accelerated virtual screening platform for drug discovery. *Nat Commun* *15*, 7761. <https://doi.org/10.1038/s41467-024-52061-7>.
19. Gentile, F., Fernandez, M., Ban, F., Ton, A.-T., Mslati, H., Perez, C.F., Leblanc, E., Yaacoub, J.C., Gleave, J., Stern, A., et al. (2021). Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus Deep Docking of 40 billion small molecules. *Chem. Sci.* *12*, 15960–15974. <https://doi.org/10.1039/D1SC05579H>.
20. Brezinova, M., Brotzakis, Z.F., Horne, R.I., Roy Chowdhury, V., Gregory, R.C., Bian, Y., González Díaz, A., Gentile, F., and Vendruscolo, M. (2025). Identification of potent high-affinity secondary nucleation inhibitors of Aβ42 aggregation from an ultra-large chemical library using deep docking. *Mol. Syst. Biol.*, 1–19.
21. Fink, E.A., Bardine, C., Gahbauer, S., Singh, I., Detomasi, T.C., White, K., Gu, S., Wan, X., Chen, J., Ary, B., et al. (2023). Large library docking for novel SARS-CoV-2 main protease non-covalent and covalent inhibitors. *Protein Science* *32*, e4712. <https://doi.org/10.1002/pro.4712>.
22. Lutzens, A., Gullberg, H., Abdurakhmanov, E., Vo, D.D., Akaberi, D., Talibov, V.O., Nekhotiaeva, N., Vangeel, L., De Jonghe, S., Jochmans, D., et al. (2022). Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J. Am. Chem. Soc.* *144*, 2905–2920. <https://doi.org/10.1021/jacs.1c08402>.

23. Gahbauer, S., Correy, G.J., Schuller, M., Ferla, M.P., Doruk, Y.U., Rachman, M., Wu, T., Diolaiti, M., Wang, S., Neitz, R.J., et al. (2023). Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2. *Proceedings of the National Academy of Sciences* 120, e2212931120. <https://doi.org/10.1073/pnas.2212931120>.
24. Singh, I., Li, F., Fink, E.A., Chau, I., Li, A., Rodriguez-Hernández, A., Glenn, I., Zapatero-Belinchón, F.J., Rodriguez, M.L., Devkota, K., et al. (2023). Structure-Based Discovery of Inhibitors of the SARS-CoV-2 Nsp14 N7-Methyltransferase. *J. Med. Chem.* 66, 7785–7803. <https://doi.org/10.1021/acs.jmedchem.2c02120>.
25. Liu, F., Mailhot, O., Glenn, I.S., Vigneron, S.F., Bassim, V., Xu, X., Fonseca-Valencia, K., Smith, M.S., Radchenko, D.S., Fraser, J.S., et al. (2025). The impact of library size and scale of testing on virtual screening. *Nat Chem Biol* 21, 1039–1045. <https://doi.org/10.1038/s41589-024-01797-w>.
26. Kaplan, A.L., Confair, D.N., Kim, K., Barros-Álvarez, X., Rodriguiz, R.M., Yang, Y., Kweon, O.S., Che, T., McCorvy, J.D., Kamber, D.N., et al. (2022). Bespoke library docking for 5-HT_{2A} receptor agonists with antidepressant activity. *Nature* 610, 582–591. <https://doi.org/10.1038/s41586-022-05258-z>.
27. Alon, A., Lyu, J., Braz, J.M., Tummino, T.A., Craik, V., O'Meara, M.J., Webb, C.M., Radchenko, D.S., Moroz, Y.S., Huang, X.-P., et al. (2021). Structures of the σ_2 receptor enable docking for bioactive ligand discovery. *Nature* 600, 759–764. <https://doi.org/10.1038/s41586-021-04175-x>.
28. Díaz-Holguín, A., Saarinen, M., Vo, D.D., Sturchio, A., Branzell, N., Cabeza de Vaca, I., Hu, H., Mitjavila-Domènech, N., Lindqvist, A., Baranczewski, P., et al. (2024). AlphaFold accelerated discovery of psychotropic agonists targeting the trace amine-associated receptor 1. *Science Advances* 10, eadn1524. <https://doi.org/10.1126/sciadv.adn1524>.
29. Tummino, T.A., Iliopoulos-Tsoutsouvas, C., Braz, J.M., O'Brien, E.S., Stein, R.M., Craik, V., Tran, N.K., Ganapathy, S., Liu, F., Shiimura, Y., et al. (2025). Virtual library docking for cannabinoid-1 receptor agonists with reduced side effects. *Nat Commun* 16, 2237. <https://doi.org/10.1038/s41467-025-57136-7>.
30. Grotzsch, K., Sadybekov, A.V., Hiller, S., Zaidi, S., Eremin, D., Le, A., Liu, Y., Smith, E.C., Iliopoulos-Tsoutsouvas, C., Thomas, J., et al. (2024). Virtual Screening of a Chemically Diverse “Superscaffold” Library Enables Ligand Discovery for a Key GPCR Target. *ACS Chem. Biol.* 19, 866–874. <https://doi.org/10.1021/acscchembio.3c00602>.
31. Stein, R.M., Kang, H.J., McCorvy, J.D., Glatfelter, G.C., Jones, A.J., Che, T., Slocum, S., Huang, X.-P., Savych, O., Moroz, Y.S., et al. (2020). Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* 579, 609–614. <https://doi.org/10.1038/s41586-020-2027-0>.
32. Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P.D., Coote, P.W., Padmanabha Das, K.M., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 580, 663–668. <https://doi.org/10.1038/s41586-020-2117-z>.

33. Fink, E.A., Xu, J., Hübner, H., Braz, J.M., Seemann, P., Avet, C., Craik, V., Weikert, D., Schmidt, M.F., Webb, C.M., et al. (2022). Structure-based discovery of nonopioid analgesics acting through the α 2A-adrenergic receptor. *Science* 377, eabn7065. <https://doi.org/10.1126/science.abn7065>.
34. Reis, J., Gorgulla, C., Massari, M., Marchese, S., Valente, S., Noce, B., Basile, L., Törner, R., Cox, H., Viennet, T., et al. (2023). Targeting ROS production through inhibition of NADPH oxidases. *Nat Chem Biol* 19, 1540–1550. <https://doi.org/10.1038/s41589-023-01457-5>.
35. Liu, F., Wu, C.-G., Tu, C.-L., Glenn, I., Meyerowitz, J., Kaplan, A.L., Lyu, J., Cheng, Z., Tarkhanova, O.O., Moroz, Y.S., et al. (2024). Large library docking identifies positive allosteric modulators of the calcium-sensing receptor. *Science* 385, eado1868. <https://doi.org/10.1126/science.ado1868>.
36. Liu, F., Kaplan, A.L., Levring, J., Einsiedel, J., Tiedt, S., Distler, K., Omattage, N.S., Kondratov, I.S., Moroz, Y.S., Pietz, H.L., et al. (2024). Structure-based discovery of CFTR potentiators and inhibitors. *Cell* 187, 3712-3725.e34. <https://doi.org/10.1016/j.cell.2024.04.046>.
37. Singh, I., Seth, A., Billesbølle, C.B., Braz, J., Rodriguiz, R.M., Roy, K., Bekele, B., Craik, V., Huang, X.-P., Boytsov, D., et al. (2023). Structure-based discovery of conformationally selective inhibitors of the serotonin transporter. *Cell* 186, 2160-2175.e17. <https://doi.org/10.1016/j.cell.2023.04.010>.
38. Vigneron, S.F., Ohno, S., Braz, J., Kim, J.Y., Kweon, O.S., Webb, C., Billesbølle, C.B., Srinivasan, K., Bhardwaj, K., Irwin, J.J., et al. (2025). Docking 14 Million Virtual Isoquinuclidines against the μ and κ Opioid Receptors Reveals Dual Antagonists–Inverse Agonists with Reduced Withdrawal Effects. *ACS Cent. Sci.* 11, 770–790. <https://doi.org/10.1021/acscentsci.5c00052>.
39. Wang, S., Wacker, D., Levit, A., Che, T., Betz, R.M., McCorvy, J.D., Venkatakrishnan, A.J., Huang, X.-P., Dror, R.O., Shoichet, B.K., et al. (2017). D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* 358, 381–386. <https://doi.org/10.1126/science.aan5468>.
40. Powers, R.A., Morandi, F., and Shoichet, B.K. (2002). Structure-based discovery of a novel, noncovalent inhibitor of AmpC β -lactamase. *Structure* 10, 1013–1023.
41. Cree, B., Bieniek, M.K., Amin, S., Kawamura, A., and Cole, D.J. (2025). Active learning driven prioritisation of compounds from on-demand libraries targeting the SARS-CoV-2 main protease. *DIGIT DISCOV.* 4, 438–450.
42. Nazarova, A.L., Sadybekov, A.V., Sadybekov, A.A., Protopopov, M., Radchenko, D.S., Moroz, Y.S., Tarkhanova, O.O., and Katritch, V. (2025). V-synthes2-the Next Generation Tool for Structure-based Virtual Screening of Giga-scale Chemical Spaces. *Research Square*, <https://doi.org/10.21203/rs.3.rs-7782723/v1>.
43. Lutten, A., Cabeza de Vaca, I., Sparring, L., Brea, J., Martínez, A.L., Kahlous, N.A., Radchenko, D.S., Moroz, Y.S., Loza, M.I., Norinder, U., et al. (2025). Rapid traversal of

- vast chemical space using machine learning-guided docking screens. *Nat Comput Sci* 5, 301–312. <https://doi.org/10.1038/s43588-025-00777-x>.
44. Kitchen, D.B., Decornez, H., Furr, J.R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949.
 45. Allen, W.J., Balias, T.E., Mukherjee, S., Brozell, S.R., Moustakas, D.T., Lang, P.T., Case, D.A., Kuntz, I.D., and Rizzo, R.C. (2015). DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* 36, 1132–1156.
 46. Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461.
 47. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., et al. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 47, 1739–1749. <https://doi.org/10.1021/jm0306430>.
 48. Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., and Banks, J.L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759.
 49. Jain, A.N. (2007). Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput. Aided Mol. Des.* 21, 281–306.
 50. Lorber, D.M., and Shoichet, B.K. (1998). Flexible ligand docking using conformational ensembles. *Protein Sci.* 7, 938–950.
 51. McGann, M. (2011). FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 51, 578–596.
 52. Depoortere, R., Boulay, D., Perrault, G., Bergis, O., Decobert, M., Françon, D., Jung, M., Simiand, J., Soubrié, P., and Scatton, B. (2003). SSR181507, A Dopamine D2 Receptor Antagonist and 5-HT1A Receptor Agonist. II: Behavioral Profile Predictive of an Atypical Antipsychotic Activity. *Neuropsychopharmacol* 28, 1889–1902. <https://doi.org/10.1038/sj.npp.1300261>.
 53. Roth, B.L., Sheffler, D.J., and Kroeze, W.K. (2004). Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* 3, 353–359. <https://doi.org/10.1038/nrd1346>.
 54. Takahashi, K., Nakagawasai, O., Nemoto, W., Odaira, T., Sakuma, W., and Tan-No, K. (2018). Antidepressant-like effect of aripiprazole via 5-HT1A, D1, and D2 receptors in the prefrontal cortex of olfactory bulbectomized mice. *Journal of Pharmacological Sciences* 137, 241–247. <https://doi.org/10.1016/j.jphs.2018.06.006>.
 55. Bailey, S.J., and Husbands, S.M. (2018). Targeting opioid receptor signaling in depression: do we need selective κ opioid receptor antagonists? *Neuronal Signal* 2, NS20170145. <https://doi.org/10.1042/NS20170145>.

56. Schmidt, D., Bernat, V., Brox, R., Tschammer, N., and Kolb, P. (2015). Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity Using Multi-Target Docking. *ACS Chem. Biol.* *10*, 715–724. <https://doi.org/10.1021/cb500577j>.
57. Apsel, B., Blair, J.A., Gonzalez, B., Nazif, T.M., Feldman, M.E., Aizenstein, B., Hoffman, R., Williams, R.L., Shokat, K.M., and Knight, Z.A. (2008). Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat Chem Biol* *4*, 691–699. <https://doi.org/10.1038/nchembio.117>.
58. Knight, Z.A., Lin, H., and Shokat, K.M. (2010). Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* *10*, 130–137. <https://doi.org/10.1038/nrc2787>.
59. Antolin, A.A., Workman, P., Mestres, J., and Al-Lazikani, B. (2016). Polypharmacology in Precision Oncology: Current Applications and Future Prospects. *Curr Pharm Des* *22*, 6935–6945. <https://doi.org/10.2174/1381612822666160923115828>.
60. Kuenzi, B.M., Remsing Rix, L.L., Stewart, P.A., Fang, B., Kinose, F., Bryant, A.T., Boyle, T.A., Koomen, J.M., Haura, E.B., and Rix, U. (2017). Polypharmacology-based ceritinib repurposing using integrated functional proteomics. *Nat Chem Biol* *13*, 1222–1231. <https://doi.org/10.1038/nchembio.2489>.
61. Sumi, N.J., Ctortocka, C., Hu, Q., Bryant, A.T., Fang, B., Remsing Rix, L.L., Ayaz, M., Kinose, F., Welsh, E.A., Eschrich, S.A., et al. (2019). Divergent Polypharmacology-Driven Cellular Activity of Structurally Similar Multi-Kinase Inhibitors through Cumulative Effects on Individual Targets. *Cell Chemical Biology* *26*, 1240-1252.e11. <https://doi.org/10.1016/j.chembiol.2019.06.003>.
62. Gentile, F., Agrawal, V., Hsing, M., Ton, A.-T., Ban, F., Norinder, U., Gleave, M.E., and Cherkasov, A. (2020). Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* *6*, 939–949.
63. Yang, Y., Yao, K., Repasky, M.P., Leswing, K., Abel, R., Shoichet, B.K., and Jerome, S.V. (2021). Efficient exploration of chemical space with docking and deep learning. *J. Chem. Theory Comput.* *17*, 7106–7119.
64. Schneider, P., Walters, W.P., Plowright, A.T., Sieroka, N., Listgarten, J., Goodnow, R.A., Fisher, J., Jansen, J.M., Duca, J.S., Rush, T.S., et al. (2020). Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov* *19*, 353–364. <https://doi.org/10.1038/s41573-019-0050-3>.
65. Walters, W.P., and Murcko, M. (2020). Assessing the impact of generative AI on medicinal chemistry. *Nat Biotechnol* *38*, 143–145. <https://doi.org/10.1038/s41587-020-0418-2>.
66. Du, Y., Jamasb, A.R., Guo, J., Fu, T., Harris, C., Wang, Y., Duan, C., Liò, P., Schwaller, P., and Blundell, T.L. (2024). Machine learning-aided generative molecular design. *Nat Mach Intell* *6*, 589–604. <https://doi.org/10.1038/s42256-024-00843-5>.
67. Catacutan, D.B., Alexander, J., Arnold, A., and Stokes, J.M. (2024). Machine learning in preclinical drug discovery. *Nat Chem Biol* *20*, 960–973. <https://doi.org/10.1038/s41589-024-01679-1>.

68. Li, Y., Pei, J., and Lai, L. (2021). Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* 12, 13664–13675. <https://doi.org/10.1039/D1SC04444C>.
69. Atance, S.R., Diez, J.V., Engkvist, O., Olsson, S., and Mercado, R. (2022). De Novo Drug Design Using Reinforcement Learning with Graph-Based Deep Generative Models. *J. Chem. Inf. Model.* 62, 4863–4872. <https://doi.org/10.1021/acs.jcim.2c00838>.
70. Iwata, H., Nakai, T., Koyama, T., Matsumoto, S., Kojima, R., and Okuno, Y. (2023). VGAE-MCTS: A New Molecular Generative Model Combining the Variational Graph Auto-Encoder and Monte Carlo Tree Search. *J. Chem. Inf. Model.* 63, 7392–7400. <https://doi.org/10.1021/acs.jcim.3c01220>.
71. Du, H., Jiang, D., Zhang, O., Wu, Z., Gao, J., Zhang, X., Wang, X., Deng, Y., Kang, Y., Li, D., et al. (2023). A flexible data-free framework for structure-based de novo drug design with reinforcement learning. *Chemical Science* 14, 12166–12181. <https://doi.org/10.1039/D3SC04091G>.
72. Roucairol, M., Georgiou, A., Cazenave, T., Prischi, F., and Pardo, O.E. (2024). DrugSynthMC: An Atom-Based Generation of Drug-like Molecules with Monte Carlo Search. *J. Chem. Inf. Model.* 64, 7097–7107. <https://doi.org/10.1021/acs.jcim.4c01451>.
73. Weller, J.A., and Rohs, R. (2024). Structure-Based Drug Design with a Deep Hierarchical Generative Model. *J. Chem. Inf. Model.* 64, 6450–6463. <https://doi.org/10.1021/acs.jcim.4c01193>.
74. Swanson, K., Liu, G., Catacutan, D.B., Arnold, A., Zou, J., and Stokes, J.M. (2024). Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. *Nat Mach Intell* 6, 338–353. <https://doi.org/10.1038/s42256-024-00809-7>.
75. Loeffler, H.H., He, J., Tibo, A., Janet, J.P., Voronov, A., Mervin, L.H., and Engkvist, O. (2024). Reinvent 4: Modern AI-driven generative molecule design. *J Cheminform* 16, 20. <https://doi.org/10.1186/s13321-024-00812-5>.
76. Wang, M., Li, S., Wang, J., Zhang, O., Du, H., Jiang, D., Wu, Z., Deng, Y., Kang, Y., Pan, P., et al. (2024). ClickGen: Directed exploration of synthesizable chemical space via modular reactions and reinforcement learning. *Nat Commun* 15, 10127. <https://doi.org/10.1038/s41467-024-54456-y>.
77. Koziarski, M., Rekesh, A., Shevchuk, D., Sloot, A. van der, Gaiński, P., Bengio, Y., Liu, C.-H., Tyers, M., and Batey, R.A. (2024). RGFN: Synthesizable Molecular Generation Using GFlowNets. Preprint at arXiv, <https://doi.org/10.48550/arXiv.2406.08506>.
78. Schneuing, A., Harris, C., Du, Y., Didi, K., Jamasb, A., Igashov, I., Du, W., Gomes, C., Blundell, T.L., Lio, P., et al. (2024). Structure-based drug design with equivariant diffusion models. *Nat Comput Sci* 4, 899–909. <https://doi.org/10.1038/s43588-024-00737-x>.
79. Zhu, Y., Li, S., Chen, J., Zhao, D., Wang, X., Li, Y., Liu, Y., Kong, Y., Zhang, B., Liu, C., et al. (2026). SynGFN: learning across chemical space with generative flow-based molecular discovery. *Nat Comput Sci* 6, 29–38. <https://doi.org/10.1038/s43588-025-00902-w>.

80. Krishnan, A., Anahtar, M.N., Valeri, J.A., Jin, W., Donghia, N.M., Sieben, L., Lutten, A., Zhang, Y., Modaresi, S.M., Hennes, A., et al. (2025). A generative deep learning approach to *de novo* antibiotic design. *Cell* 188, 5962-5979.e22.
<https://doi.org/10.1016/j.cell.2025.07.033>.
81. Eisenhuth, P., Liessmann, F., Moretti, R., and Meiler, J. (2025). Ultra-large library screening with an evolutionary algorithm in Rosetta (REvoLd). *Commun Chem* 8, 335.
<https://doi.org/10.1038/s42004-025-01758-x>.
82. Cretu, M., Harris, C., Igashov, I., Schneuing, A., Segler, M., Correia, B., Roy, J., Bengio, E., and Liò, P. (2025). SynFlowNet: Design of Diverse and Novel Molecules with Synthesis Constraints. Preprint at arXiv, <https://doi.org/10.48550/arXiv.2405.01155>
<https://doi.org/10.48550/arXiv.2405.01155>.
83. Powers, A.S., Lu, T., Koodli, R.V., Xu, M., Gu, S., Karelina, M., and Dror, R.O. (2025). MedSAGE: Bridging Generative AI and Medicinal Chemistry for Structure-Based Design of Small Molecule Drugs. Preprint at bioRxiv, <https://doi.org/10.1101/2025.05.10.653107>
<https://doi.org/10.1101/2025.05.10.653107>.
84. Balias, T.E., Tan, Y.S., and Chakrabarti, M. (2024). DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *J. Comput. Chem.* 45, 47–63.
85. Day, P.N., Jensen, J.H., Gordon, M.S., Webb, S.P., Stevens, W.J., Krauss, M., Garmer, D., Basch, H., and Cohen, D. (1996). An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* 105, 1968–1986.
86. Stein, R.M., Yang, Y., Balias, T.E., O'Meara, M.J., Lyu, J., Young, J., Tang, K., Shoichet, B.K., and Irwin, J.J. (2021). Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model.* 61, 699–714.
87. Chen, P., Ke, Y., Lu, Y., Du, Y., Li, J., Yan, H., Zhao, H., Zhou, Y., and Yang, Y. (2019). DLIGAND2: an improved knowledge-based energy function for protein–ligand interactions using the distance-scaled, finite, ideal-gas reference state. *J. Cheminform.* 11, 52.
88. Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., and Steffenel, L.A. (2022). Machine-learning methods for ligand–protein molecular docking. *Drug Discovery Today* 27, 151–164.
89. Eberhardt, J., Santos-Martins, D., Tillack, A.F., and Forli, S. (2021). AutoDock Vina 1.2. 0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* 61, 3891–3898.
90. Shen, C., Zhang, X., Deng, Y., Gao, J., Wang, D., Xu, L., Pan, P., Hou, T., and Kang, Y. (2022). Boosting protein–ligand binding pose prediction and virtual screening based on residue–atom distance likelihood potential and graph transformer. *J. Med. Chem.* 65, 10691–10706.
91. Lyu, J., Irwin, J.J., and Shoichet, B.K. (2023). Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* 19, 712–718.

92. Wang, W.W., Gallo, L., Jadhav, A., Hawkins, R., and Parker, C.G. (2020). The Druggability of Solute Carriers. *J. Med. Chem.* 63, 3834–3867. <https://doi.org/10.1021/acs.jmedchem.9b01237>.
93. Galetin, A., Brouwer, K.L.R., Tweedie, D., Yoshida, K., Sjöstedt, N., Aleksunes, L., Chu, X., Evers, R., Hafey, M.J., Lai, Y., et al. (2024). Membrane transporters in drug development and as determinants of precision medicine. *Nat Rev Drug Discov* 23, 255–280. <https://doi.org/10.1038/s41573-023-00877-1>.
94. Girardi, E., César-Razquin, A., Lindinger, S., Papakostas, K., Konecka, J., Hemmerich, J., Kicking, S., Kartnig, F., Gürtl, B., Klavins, K., et al. (2020). A widespread role for SLC transmembrane transporters in resistance to cytotoxic drugs. *Nat Chem Biol* 16, 469–478. <https://doi.org/10.1038/s41589-020-0483-3>.
95. Chiu, T.-Y., Lazar, D.C., Wang, W.W., Wozniak, J.M., Jadhav, A.M., Li, W., Gazaniga, N., Theofilopoulos, A.N., Teijaro, J.R., and Parker, C.G. (2024). Chemoproteomic development of SLC15A4 inhibitors with anti-inflammatory activity. *Nat. Chem. Biol.* 20, 1000–1011.
96. Boeszoermenyi, A., Bernaleau, L., Chen, X., Kartnig, F., Xie, M., Zhang, H., Zhang, S., Delacrétaz, M., Koren, A., Hopp, A.-K., et al. (2023). A conformation-locking inhibitor of SLC15A4 with TASL proteostatic anti-inflammatory activity. *Nat Commun* 14, 6626. <https://doi.org/10.1038/s41467-023-42070-3>.
97. Drew, D., and Boudker, O. (2024). Ion and lipid orchestration of secondary active transport. *Nature* 626, 963–974.
98. Im, D., Jormakka, M., Juge, N., Kishikawa, J., Kato, T., Sugita, Y., Noda, T., Uemura, T., Shiimura, Y., Miyaji, T., et al. (2024). Neurotransmitter recognition by human vesicular monoamine transporter 2. *Nat Commun* 15, 7661. <https://doi.org/10.1038/s41467-024-51960-z>.
99. Garib Singh, R.-A.A., and Schlessinger, A. (2019). Advances and challenges in rational drug design for SLCs. *Trends Pharmacol. Sci.* 40, 790–800.
100. Schlessinger, A., Zatorski, N., Hutchinson, K., and Colas, C. (2023). Targeting SLC transporters: small molecules as modulators and therapeutic opportunities. *Trends Biochem. Sci* 48, 801–814.
101. Zhang, Y., Dai, F., Chen, N., Zhou, D., Lee, C.-H., Song, C., Zhang, Y., and Zhang, Z. (2024). Structural insights into VACHT neurotransmitter recognition and inhibition. *Cell Res.* 34, 665–668.
102. Zhang, Y., and Zhang, Z. (2026). Structural basis of VACHT inhibition by spiroindolines and alkylsulfones. *Structure* 34, 123-132. e4.
103. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.

104. Dalton, M.P., Cheng, M.H., Bahar, I., and Coleman, J. (2024). Structural mechanisms for VMAT2 inhibition by tetrabenazine. *Biophys. J.* 123, 116a.
105. Wei, F., Liu, H., Zhang, W., Wang, J., and Zhang, Y. (2025). Drug inhibition and substrate transport mechanisms of human VMAT2. *Nat. Commun.* 16, 323.
106. Wu, D., Chen, Q., Yu, Z., Huang, B., Zhao, J., Wang, Y., Su, J., Zhou, F., Yan, R., Li, N., et al. (2024). Transport and inhibition mechanisms of human VMAT2. *Nature* 626, 427–434. <https://doi.org/10.1038/s41586-023-06926-4>.
107. Zheng, G., Dwoskin, L.P., and Crooks, P.A. (2006). Vesicular monoamine transporter 2: role as a novel target for drug development. *AAPS J.* 8, 78.
108. Gamez, J., Calopa, M., Muñoz, E., Ferré, A., Huertas, O., McAllister, K., Reig, N., Scart-Grès, C., Insa, R., and Kulisevsky, J. (2023). A proof-of-concept study with SOM3355 (bevantolol hydrochloride) for reducing chorea in Huntington's disease. *Br. J. Clin. Pharmacol.* 89, 1656–1664.
109. Rosenthal, L.S., Farag, M., Aziz, N.A., and Bang, J. (2025). Vesicular monoamine transport inhibitors: current uses and future directions. *The Lancet* 406, 650–664.
110. Ma, Q., Ma, K., Dong, Y., Meng, Y., Zhao, J., Li, R., Bai, Q., Wu, D., Jiang, D., Sun, J., et al. (2025). Binding mechanism and antagonism of the vesicular acetylcholine transporter VACHT. *Nat Struct Mol Biol* 32, 818–827. <https://doi.org/10.1038/s41594-024-01462-9>.
111. O'Grady, G.L., Verschuuren, C., Yuen, M., Webster, R., Menezes, M., Fock, J.M., Pride, N., Best, H.A., Benavides Damm, T., Turner, C., et al. (2016). Variants in SLC18A3, vesicular acetylcholine transporter, cause congenital myasthenic syndrome. *Neurology* 87, 1442–1448. <https://doi.org/10.1212/WNL.00000000000003179>.
112. Ogawa, K., and Shiba, K. (2018). In vivo and in vitro characteristics of radiolabeled vesamicol analogs as the vesicular acetylcholine transporter imaging agents. *Contrast Media Mol. Imaging* 2018, 4535476.
113. Pidathala, S., Liao, S., Dai, Y., Li, X., Long, C., Chang, C.-L., Zhang, Z., and Lee, C.-H. (2023). Mechanisms of neurotransmitter transport and drug inhibition in human VMAT2. *Nature* 623, 1086–1092.
114. Barthel, C., Sorger, D., Deuther-Conrad, W., Scheunemann, M., Schweiger, S., Jäckel, P., Roghani, A., Steinbach, J., Schüürmann, G., Sabri, O., et al. (2015). New systematically modified vesamicol analogs and their affinity and selectivity for the vesicular acetylcholine transporter – A critical examination of the lead structure. *European Journal of Medicinal Chemistry* 100, 50–67. <https://doi.org/10.1016/j.ejmech.2015.05.033>.
115. Lindemann, M., Deuther-Conrad, W., Moldovan, R., Sekhar, K.V.G.C., Brust, P., and Wenzel, B. (2017). Do spiroindolines have the potential to replace vesamicol as lead compound for the development of radioligands targeting the vesicular acetylcholine transporter? *Bioorganic & Medicinal Chemistry* 25, 5107–5113. <https://doi.org/10.1016/j.bmc.2017.03.028>.

116. Bemis, G.W., and Murcko, M.A. (1996). The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39, 2887–2893.
117. Tang, S., Chen, R., Lin, M., Lin, Q., Zhu, Y., Ding, J., Hu, H., Ling, M., and Wu, J. (2022). Accelerating autodock vina with gpus. *Molecules* 27, 3041.
118. Ding, J., Tang, S., Mei, Z., Wang, L., Huang, Q., Hu, H., Ling, M., and Wu, J. (2023). Vina-GPU 2.0: Further Accelerating AutoDock Vina and Its Derivatives with Graphics Processing Units. *J. Chem. Inf. Model.* 63, 1982–1998. <https://doi.org/10.1021/acs.jcim.2c01504>.
119. Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V.R., Getz, N., Portnoi, T., Roy, J., Stark, H., et al. (2025). Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. Preprint at bioRxiv, <https://doi.org/10.1101/2025.06.14.659707> <https://doi.org/10.1101/2025.06.14.659707>.
120. Hall, B.W., Tummino, T.A., Tang, K., Mailhot, O., Castanon, M., Irwin, J.J., and Shoichet, B.K. (2025). A database for large-scale docking and experimental results. *J. Chem. Inf. Model.* 65, 4458–4467.
121. Durant, G., Boyles, F., Birchall, K., and Deane, C.M. (2024). The future of machine learning for small-molecule drug discovery will be driven by data. *Nat. Comput. Sci.* 4, 735–743.
122. Kramer, C., Chodera, J., Damm-Ganamet, K.L., Gilson, M.K., Günther, J., Lessel, U., Lewis, R.A., Mobley, D., Nittinger, E., Pecina, A., et al. (2025). The Need for Continuing Blinded Pose- and Activity Prediction Benchmarks. *J. Chem. Inf. Model.* 65, 2180–2190. <https://doi.org/10.1021/acs.jcim.4c02296>.
123. Rodriguez, D.C.P., Weber, K.C., Sundberg, B., and Glasgow, A. (2024). MAGPIE: An interactive tool for visualizing and analyzing protein–ligand interactions. *Protein Science* 33, e5027. <https://doi.org/10.1002/pro.5027>.
124. Marchand, A., Buckley, S., Schneuing, A., Pacesa, M., Elia, M., Gainza, P., Elizandrova, E., Neeser, R.M., Lee, P.-W., Reymond, L., et al. (2025). Targeting protein–ligand neosurfaces with a generalizable deep learning tool. *Nature* 639, 522–531. <https://doi.org/10.1038/s41586-024-08435-4>.
125. Masters, M.R., Mahmoud, A.H., and Lill, M.A. (2025). Investigating whether deep learning models for co-folding learn the physics of protein–ligand interactions. *Nat. Commun.* 16, 8854.
126. Tingle, B.I., and Irwin, J.J. (2023). Large-scale docking in the cloud. *J. Chem. Inf. Model.* 63, 2735–2741.
127. Coleman, R.G., Carchia, M., Sterling, T., Irwin, J.J., and Shoichet, B.K. (2013). Ligand pose and orientational sampling in molecular docking. *PLoS One* 8, e75992.
128. Riniker, S., and Landrum, G.A. (2015). Better informed distance geometry: using what we know to improve conformation generation. *J. Chem. Inf. Model.* 55, 2562–2574.

129. Hawkins, P.C., Skillman, A.G., Warren, G.L., Ellingson, B.A., and Stahl, M.T. (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* *50*, 572–584.
130. Wei, B.Q., Baase, W.A., Weaver, L.H., Matthews, B.W., and Shoichet, B.K. (2002). A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* *322*, 339–355.
131. Mysinger, M.M., and Shoichet, B.K. (2010). Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* *50*, 1561–1573.
132. Sterling, T., and Irwin, J.J. (2015). ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.* *55*, 2324–2337.
133. Meng, E.C., Shoichet, B.K., and Kuntz, I.D. (1992). Automated docking with grid-based energy evaluation. *J. Comput. Chem.* *13*, 505–524.
134. Gallagher, K., and Sharp, K. (1998). Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* *75*, 769–776.
135. Tan, Y.S., Chakrabarti, M., Stein, R.M., Prentis, L.E., Rizzo, R.C., Kurtzman, T., Fischer, M., and Balius, T.E. (2025). Development of receptor desolvation scoring and covalent sampling in DOCK 6: Methods evaluated on a RAS test set. *J. Chem. Inf. Model.* *65*, 722–748.
136. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* *529*, 484–489. <https://doi.org/10.1038/nature16961>.
137. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* *25*, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
138. Heid, E., Greenman, K.P., Chung, Y., Li, S.-C., Graff, D.E., Vermeire, F.H., Wu, H., Green, W.H., and McGill, C.J. (2024). Chemprop: A Machine Learning Package for Chemical Property Prediction. *J. Chem. Inf. Model.* *64*, 9–17. <https://doi.org/10.1021/acs.jcim.3c01250>.
139. Fassio, A.V., Shub, L., Ponzoni, L., McKinley, J., O'Meara, M.J., Ferreira, R.S., Keiser, M.J., and de Melo Minardi, R.C. (2022). Prioritizing virtual screening with interpretable interaction fingerprints. *J. Chem. Inf. Model.* *62*, 4300–4318.
140. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 12–21.

141. Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., and Ke, G. (2023). Uni-mol: A universal 3d molecular representation learning framework. In.
142. Sayle, R.A., Batista, J., and Grant, J.A. (2013). Efficient maximum common subgraph (MCS) searching of large chemical databases. *J. Cheminform.* 5, O15.
143. Guo, W., Wang, M., and Chen, L. (2022). A co-expression vector for baculovirus-mediated protein expression in mammalian cells. *Biochem. Biophys. Res. Commun.* 594, 69–73.
144. Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G., and Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* 22, 575–582.
145. Goehring, A., Lee, C.-H., Wang, K.H., Michel, J.C., Claxton, D.P., Bacongus, I., Althoff, T., Fischer, S., Garcia, K.C., and Gouaux, E. (2014). Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* 9, 2574–2585.
146. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296.
147. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* 17, 1214–1221. <https://doi.org/10.1038/s41592-020-00990-8>.
148. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* 16, 1153–1160.
149. Wang, N., Jiang, X., Zhang, S., Zhu, A., Yuan, Y., Xu, H., Lei, J., and Yan, C. (2021). Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. *Cell* 184, 370–383. e13.
150. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25.
151. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 74, 531–544.
152. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
153. Jakalian, A., Bush, B.L., Jack, D.B., and Bayly, C.I. (2000). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21, 132–146.
154. Jakalian, A., Jack, D.B., and Bayly, C.I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 23, 1623–1641.

