



Published in final edited form as:

Nature. 2022 January ; 601(7893): 452–459. doi:10.1038/s41586-021-04220-9.

Synthon-based ligand discovery in virtual libraries of over 11 billion compounds

Arman A. Sadybekov^{1,2,§}, Anastasiia V. Sadybekov^{1,2,§}, Yongfeng Liu^{3,4,§}, Christos Iliopoulos-Tsoutsouvas^{5,§}, Xi-Ping Huang^{3,6}, Julie Pickett^{3,6}, Blake Houser², Nilkanth Patel¹, Ngan K. Tran⁵, Fei Tong⁵, Nikolai Zvonok⁵, Manish K Jain³, Olena Savych⁷, Dmytro S. Radchenko^{7,8}, Spyros P. Nikas⁵, Nicos A. Petasis², Yurii S. Moroz^{8,9}, Bryan L. Roth^{3,4,6,*}, Alexandros Makriyannis^{5,10,*}, Vsevolod Katritch^{1,2,*}

¹Department of Quantitative and Computational Biology, Bridge Institute, USC Michelson Center for Convergent Biosciences, University of Southern California, Los Angeles, CA 90089, USA

²Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

³Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

⁴Division of Chemical Biology and Medicinal Chemistry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

⁵Center for Drug Discovery and Department of Pharmaceutical Sciences, Northeastern University, Boston, MA 02115, USA

⁶National Institute of Mental Health Psychoactive Drug Screening Program, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

⁷Enamine Ltd, 78 Chervonotkatska Street, 02094, Ukraine

⁸Taras Shevchenko National University of Kyiv, Volodymyrska Street 60, Kyiv 01601, Ukraine

⁹Chemspace LLC, 85 Chervonotkatska Street, 02094, Ukraine

¹⁰Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA

SUMMARY

*Corresponding authors: Bryan L. Roth (bryan_roth@med.unc.edu), Alexandros Makriyannis (a.makriyannis@northeastern.edu), Vsevolod Katritch (katritch@usc.edu).

§These authors contributed to the study equally

Author contributions: A.A.S. and A.V.S. developed V-SYNTHES algorithms, performed calculations, and written the first draft of the manuscript; B.H. and N.A.P. performed calculations and compound selection for ROCK1; Y.L., M.K.J., J.P., and X.P.H. performed functional and selectivity assays; C.I.T., N.K.T., F.T., N.Z., and S.P.N. performed binding assays; N.P. performed full VLS on Google Cloud; O. S., D.S.R., and Y.S.M. developed REAL Space library and performed compound synthesis; B.L.R. supervised the functional and selectivity assays; A.M. supervised binding assays for CB₁ and CB₂; V.K. conceived the study and supervised all its computational aspects. All authors contributed to writing and editing the manuscript.

Competing Interest: A.A.S and V.K. filed a provisional patent on V-SYNTHES method.

Data and Code Availability: Chemical structures, synthetic methods, detailed results of biochemical characterization are presented in extended data and supplementary information, available online at <https://...> V-SYNTHES scripts and example files have been deposited to <https://github.com/katritchlab/V-SYNTHES>.

Structure-based virtual ligand screening is emerging as a key paradigm for early drug discovery owing to the availability of high-resolution target structures¹⁻⁴ and ultra-large libraries of virtual compounds^{5,6}. However, to keep pace with the explosive growth of virtual libraries, such as REAdily AvaiLable for synthesis (REAL) combinatorial libraries⁷, new approaches to compound screening are needed^{8,9}. Here we introduce a novel synthon-based approach, V-SYNTHES, to perform hierarchical structure-based screening of REAL library of more than 11 billion compounds. V-SYNTHES first identifies the best synthon-scaffold combinations as seeds suitable for further growth, and then iteratively elaborates these seeds to select complete molecules with the best docking scores. This hierarchical combinatorial approach allows rapid detection of the best-scoring compounds in the Giga-scale chemical space while performing docking of only a small fraction (<0.1%) of the library compounds. Chemical synthesis and experimental testing of cannabinoid antagonists predicted by V-SYNTHES demonstrated 33% hit rate, including 14 submicromolar ligands, dramatically improving over a standard virtual screening of the Enamine REAL diversity subset, which required ~100 times more computational resources. Synthesis of selected analogs of the best hits further improved potencies and affinities (best K_i = 0.6 nM) and CB₂/CB₁ selectivity (50-100 fold). V-SYNTHES was also tested on a kinase target, ROCK1, further supporting its utility for lead discovery. The approach is easily scalable for the rapid growth of combinatorial libraries and potentially adaptable for any docking algorithm.

Keywords

docking; structure-based drug discovery; virtual ligand screening; ultra-large; giga-scale chemical libraries; synthon; cannabinoid receptors

Introduction

Standard libraries for high-throughput (HTS)¹⁰ and virtual ligand screening (VLS)¹¹⁻¹³ have been historically limited to <10 Million available compounds, which is a minute fraction of the enormous chemical space, estimated as 10^{20} to 10^{60} drug-like compounds^{14,15}. This limitation of standard HTS and VLS slows the pace of drug discovery, usually yielding initial hits with modest affinities, poor selectivity, and ADMET profiles, which require elaborate multistep optimization to gain lead- and drug-like candidate properties. Recently, ultra-large libraries of more than 100 Million REAdily AccessibLe (REAL) compounds have been developed and employed in docking-based VLS, yielding high-quality hits for lead discovery^{5,6}. The Enamine REAL library, now enumerating >1.4 Billion, and its REAL Space extension comprising >10 Billion drug-like compounds, take advantage of modular parallel synthesis with a large set of optimized reactions and building blocks (“synthons”)⁶. This makes the synthesis of potential hit compounds fast (less than 4-6 weeks), reliable (>80% success rate), and affordable.

The modular nature of REAL libraries supports their further rapid growth way beyond 10 Billion drug-like compounds¹⁶. With increasing library sizes, however, the computational time and cost of docking-based VLS itself become the next bottleneck in screening, even with massively parallel cloud computing capacities. For example, docking of 10 Billion compounds at a standard rate of 10 sec/compound would take >3000 years on a single

CPU core, or cost >\$800,000 on a computing cloud. The ability to dramatically reduce the computational burden of VLS, without compromising the accuracy of docking or losing the best hit compounds would remove this bottleneck and assure broad accessibility of Giga-scale screening. Recently, an iteration of docking and machine learning steps⁹, or stepwise filtering of the whole enumerated library using docking algorithms of increasing accuracy⁸ were suggested to tackle ultra-large libraries of 138M and 1.4B compounds, respectively. However, these methods still require vast computational resources that scale linearly with the growing number of compounds.

Here we present the **Virtual SYNThon Hierarchical Enumeration Screening (V-SYNTHES)** approach that takes full advantage of the modular building block organization of the Enamine REAL Space, does not need full enumeration of the library and requires thousands of times less computational resources than standard VLS without compromising docking accuracy at any step. Moreover, the algorithm cost scales linearly with the number of synthons, or as square or cubic root of the whole library size ($O(N^{1/2})$ and $O(N^{1/3})$ for 2-component and 3-component reactions respectively). Such performance of V-SYNTHES relies on the initial docking of a pre-built set of the fragment-like compounds representing all of the library reaction scaffolds and corresponding synthons. The best selected scaffold-synthon combinations are then enumerated, and the resulting focused library is docked again to select fully elaborated hits. These iterations help to focus on a small fraction (<0.1%) of the best synthons, thus drastically reducing the combinatorial chemical space for docking.

The approach was applied here to cannabinoid receptors, which are class A G-protein coupled receptors (GPCRs), and key targets in drug discovery for inflammatory, neurodegenerative diseases, and cancer¹⁷⁻¹⁹. V-SYNTHES allowed to speed up prospective screening of 11B REAL Space library >5000 fold by iteratively docking only ~2 million full compounds. Moreover, experimental validation showed that V-SYNTHES doubled the success rate in the discovery of CB hits as compared to a standard VLS screen of the REAL diversity subset of 115 Million compounds (33% vs. 15%). Similarly, V-SYNTHES application to the kinase target ROCK1 yielded 28.5% hit rate, including ligands with nanomolar affinity and potency. The new approach provides a practical alternative for fast screening of growing Giga-scale modular virtual libraries, helping to identify leads suitable for fast optimization in the same REAL Space.

Results

The REAL Space virtual library

The V-SYNTHES approach has been implemented for the REAL Space virtual library that comprised more than 11 Billion readily accessible compounds based on optimized one-pot parallel synthesis (Enamine), involving 121 reaction protocols and 75,000 unique reagents. The reaction protocols include single and multistep procedures involving two (102 reaction protocols) or three (17 reaction protocols) starting reagents. In this work, we used only 2-component and 3-component reactions yielding ~500 Million and ~10.5 Billion compounds respectively. V-SYNTHES approach can be easily expanded to 4- and more component reactions when they become a substantial part of REAL Space. Each reaction/scaffold in

the library is presented in the form of a Markush scheme with two or more R-groups representing “synthons”^{7,20}.

High diversity of the REAL space is achieved through utilizing diverse sets of starting reagents. Average numbers of starting reagents per protocol are the following: for 2-reagent reactions, 3,344 (reagent 1) and 2,068 (reagent 2); for 3-reagent reactions, 939 (reagent 1), 1,308 (reagent 2), and 1,389 (reagent 3). The modular design of the library based on well-established and optimized reactions and an automated one-pot parallel synthesis approach allows fast synthesis (less than 4-6 weeks), with a high success rate (>80%) and guaranteed high purity (>90%).

V-SYNTHES screening approach

The V-SYNTHES approach involves iterative steps of library preparation, enumeration, docking, and hit selection as outlined in Figure 1 (see also more details in Methods). In the preparatory **Step 1**, we generate a library of fragment-like compounds representing all possible scaffold-synthon combinations for all reactions in the whole Enamine REAL Space, which we will refer to as a “Minimal Enumeration Library” (**MEL**). The MEL compounds are built from the reaction scaffolds, enumerated with the corresponding synthons at one of its R-positions, while the other R-position(s) are being “capped” with a special minimal synthon according to the reaction specified for this R-position (Figure 1). This capping, usually containing methyl or phenyl moieties, is needed to convert the reactive groups of the scaffold into a chemical form that corresponds to the full compounds (e.g. primary amine into methyl-amide or secondary amine), to better match the binding properties of the full compounds. Since only one of the R groups is fully enumerated, and others are just systematically “capped”, the MEL library size is approximately equal to the number of synthons in the REAL Space, i.e. only about 600K compounds. This MEL preparation step is performed once for the REAL Space library and does not depend on the target receptor.

In **Step 2**, the MEL compounds are docked to the target receptor using energy-based docking of the flexible ligand. The results of docking, including predicted binding scores and ligand-receptor interaction information, typically for a few thousand top-scoring compounds, are then used to select the most promising fragments for the next enumeration. The selection is also filtered for diversity, including a rule that a single reaction cannot contribute more than 20% of the selection.

Step 3 involves the iterative enumeration and docking of the best MEL compounds selected in Step 2. On each iteration, the compounds are enumerated so that one of the capped R groups is replaced by a full range of corresponding synthons from the library. For example, for two-component reactions with only two R groups, a single Step 3 iteration completes the molecule, representing a full compound from the REAL Space. For three- and more component reactions, two and more iterations are performed, replacing one by one the minimal caps with real R group synthons. Thus, each “hit” MEL compound selected in the previous iteration step is combinatorially “grown”, resulting in fully enumerated compounds from the REAL Space.

Finally, **Step 4** performs the docking screen on the final enumerated subset of the library. The several thousands of top-ranked VLS hits undergo postprocessing filtering for PAINS²¹, physical-chemical properties, drug-likeness, novelty, and chemical diversity to select a final limited set (typically 50-100) compounds for synthesis and experimental testing.

The premise of this approach is to enrich the MEL library on Step 2 - and then each subsequent iteration library - with Scaffold-Synthon combinations that have high binding scores in the pocket and are suitable for further enumeration. Because of the modular combinatorial nature of the REAL Space library, narrowing down the most promising scaffolds-synthon combinations dramatically reduces the enumerated chemical space for docking, e.g. from 11B to 2M compounds in our case.

Structure-guided selection of fragments

Selection of synthons in Step 2, if based solely on binding scores, can already bring substantial library enrichment, with an estimated up to 40 enrichment of high-scoring compounds in the final iteration library than in the random subset of the full library (see Extended Data Figure 1). At the same time, we found that the performance of the iterative approach can be further improved by taking into account docking poses of the compounds, and specifically, positions of the minimal capping R-group. Thus, docking of the fragments into a binding pocket can result in two conceptually different outcomes. The first, “productive” outcome, is when the minimal capping group of the docked MEL ligand is positioned in the pocket in such a way that it can be replaced by real, bulkier synthons from the library upon the next step of enumeration. This requires the cap to be pointing toward the unoccupied part of the pocket and not being blocked by the pocket residues. A second, “non-productive” outcome is when the minimal cap at one of the R-positions is directly pointing towards the residues at the dead-end sub-pocket, where it does not have space to grow. Another non-productive situation is when the capping R-group is pointing outside of the pocket, where useful contacts are much less likely. To select productive hits, we used an automated procedure that checks the distance from the cap atoms to selected (dummy) atoms at the dead-end subpockets. The corresponding rules in implementation for the CB₂ receptor are described in Extended Data Figure 2. The docked MEL compounds for which their cap atoms approached the “dead-end” residues closer than 4Å were excluded from further consideration even if they had high-ranked binding scores.

Screening CB receptors with V-SYNTHES

The V-SYNTHES approach was then applied to screen 11 Billion REAL Space compounds using recently solved representative CB₂R structure in complex with an antagonist (PDB:5ZTY) as a template²². We performed separate screening for 2-component and 3-component reactions of the library, representing ~500M and ~10.5B virtual compounds. Note that V-SYNTHES required docking of just 1M and 0.5M compounds respectively for these libraries in the last enumeration step, reducing the computational cost of screening more than 5000-fold.

To computationally benchmark the performance of V-SYNTHES vs. a standard VLS procedure, we also generated randomized 1M and 0.5M compounds subsets from the same

2-component and 3-component REAL Space and assessed them in standard VLS using the same receptor model and same docking parameters. Note, that the full 11B REAL Space library is not amenable to standard VLS (see Introduction). Figure 2 compares the screening performance of V-SYNTHES with standard VLS over the range of docking score thresholds. The results show that V-SYNTHES detected many more high-scoring compounds with much better scores than standard VLS that involved docking of the same number of compounds. Thus, the best 2-component compound identified by V-SYNTHES scored 7 kJ/mol better than the very best hit from the standard VLS; the difference was 6.5 kJ/mol for 3-component compounds. Moreover, in the 2-component REAL space V-SYNTHES identified 84 compounds with binding scores that were better than the very best compound from standard VLS; this number was 136 for the 3-component space.

To systematically characterize the enrichment for high-scoring compounds in the final step of V-SYNTHES vs. a random subset of the whole library, we introduced the Enrichment Factor. At a given docking score threshold, the Enrichment Factor is calculated as a ratio of “number of candidate hits” detected in the V-SYNTHES final step enumerated library vs. a random subset of REAL Space with the same number of compounds, as shown in Figure 2c,d.

Note that at -30 kJ/mol binding score threshold, V-SYNTHES already yields ~ 40 - 50 -fold higher number of “potential hits” from 2-component ($>10,000$ hits) and 3-component space ($>5,000$ hits), compared to standard VLS. This enrichment further increases for more restrictive thresholds, reflecting the V-SYNTHES focus on the iterative selection of the very best-scoring compounds. One relevant way of measuring the enrichment factor is to set the docking score threshold so that it selects 100 top-scoring compounds (which we refer to as EF_{100}), where 100 is a typical number of compounds selected in VLS campaigns for synthesis and experimental testing. For the 2-component reaction, this enrichment factor was estimated as $EF_{100} = 250$. This is approaching a theoretical limit of “ideal enrichment” ~ 500 , which would be achievable if all possible hits from the full chemical space of 500M compounds were present in the 1M compound final enumerated library. For the 3-component reactions, the $EF_{100} = 460$ is even higher and sufficient for high practical utility, though further from the theoretical limit of 20,000.

The Enrichment factor evaluation did not take into account computational efforts for the initial docking of MEL compounds (and intermediate library for 3-component). However, these initial steps add only limited computational costs to V-SYNTHES screens ($\sim 20\%$ for 2-component and 35% for 3-component), because smaller fragment-like compounds in the MEL library dock much faster on average than the larger and more flexible compounds. Considering the full computational cost at all the iterative steps, the acceleration of V-SYNTHES as compared to standard screening for identification of the 100 top candidate hits at the same score threshold thus can be evaluated as ~ 200 fold for 2-component and 300-fold for 3-component compounds in the current benchmark.

Selection and synthesis of candidate hits

To select the best V-SYNTHES hits for chemical synthesis and *in-vitro* testing at CB receptors, we employed a standard post-processing procedure to the top-ranking 5000

candidate hits, which included (i) filtering out compounds with potential PAINS properties and low drug-likeness, (ii) filtering out compounds with high similarity to known CB₁/CB₂ ligands in ChEMBL, (iii) redocking initial hits at a higher docking effort, (iv) clustering and selection of a limited number of best compounds from each cluster to maintain higher diversity of the final set. The final selected set included 80 compounds, of which 60 were synthesized with >90% purity and delivered by Enamine in less than 5 weeks. The details of this selection procedure can be found in Extended Data Figure 3. The list of all synthesized compounds from V-SYNTHES screening is shown in SI Table S1, and details of compound synthesis and quality control in SI Methods and Source Data).

Characterization of new CB ligands

Initial functional characterization of 60 novel candidate ligands predicted by V-SYNTHES identified 21 compounds with antagonist activity (>40% inhibition at 10 μM concentration) at human CB₁, CB₂, or both in the β-arrestin recruitment Tango assay (SI Figures S1 and S2). Three compounds, **673**, **505**, and **599**, showed weak partial CB₂ agonism at 10 μM and or 3 μM, though also behaved like antagonists in the antagonism assays. The primary hits were tested for their antagonist potency in full 16-point dose-response assays at CB₁ and CB₂, in the presence of a fixed 100 nM concentration (EC₈₀) of the dual CB₁/CB₂ CP55,940 agonist that submaximally activates the receptors (Extended Data Figure 4). Among the 60 compounds predicted by V-SYNTHES, the Tango assays identified 21 hits with functional K_i values better than 10 μM, including 21 antagonists for CB₁ and 20 antagonists for CB₂ (see Extended Data Table 1 and Figure 3). This constitutes a high 33% hit rate for both receptors, on the high end of the range observed in prospective screening for GPCRs⁴. Among identified hit compounds, 14 showed sub-micromolar functional K_i values as antagonists at the CB₁ receptor and 3 compounds at the CB₂ receptor. The same 60 compounds were also tested in radioligand binding assays with human CB₂ and rat CB₁ receptors and [³H]CP-55,940 as the radioligand. Of these, 9 compounds had affinities (K_i) better than 10 μM to the CB₁ receptor and 16 compounds with affinities better than 10 μM to CB₂ receptor (Extended Data Table 1, Extended Data Figure 5).

To assess the broad off-target selectivity, the best compounds, **523**, **610**, and **673**, were also tested at 10 μM concentration in GPCRome-Tango assays with a panel of more than 300 receptors²³ (Extended Data Figure 6). The panel shows only a few (3-5) potential off-targets, with only negligible off-target activities in the follow-up dose-response assays.

Molecular determinants of the hits

Experimentally identified hit compounds showed a broad diversity in their chemical structures (Figure 3b-g), representing novel scaffolds with Tanimoto distance > 0.3 from known CB₁ and CB₂ ligands found in ChEMBL²⁴ (pAct > 5.0). The best hit compounds were predicted to largely fill the receptor orthosteric pocket, similar to antagonist AM10257 that was co-crystallized with CB₂ receptor²² (**Figure 5**). Best hit compounds occupy all three subpockets of the CB₂ binding pocket, where benzene ring (Subpocket 1), 5-hydroxypentyl chain (Subpocket 2), and adamantyl group (Subpocket 3) of AM10257 are bound in the crystal structure of the receptor. Like in AM10257, these interactions suggest antagonistic profiles for our hit compounds, as compared to the recently solved Cryo-EM

structure of CB₂ receptor with agonist WIN 55,212-2, which shows that agonist molecules avoid interaction with Subpocket 1 W194, F117, and W258 side chains²⁵. Subpocket 1 preferably binds aromatic ring, however, two hit compounds (**505** and **523**) fill it with a non-aromatic ring and one compound with an aliphatic substituent (**681**). Interestingly, while most previously known CB₁/CB₂ ligands, including AM10257 and THC analogs have an aliphatic moiety in Subpocket 2, our hits have more bulky cyclic groups, while compound **505** avoids this pocket altogether. Notably, while lipophilicity of CB receptor pockets represents a challenge for developing high-affinity drug-like ligands, all the V-SYNTHES derived hits have cLogP<5 and are smaller than 500 DA.

Comparison to standard VLS

In parallel to the V-SYNTHES screen, we performed a standard ultra-large VLS for a representative 115 Million compound diversity subset of the Enamine REAL library, using the same receptor model and the same parameters of the docking algorithm. As a result of this standard full-scale screening, 97 predicted hits were selected, synthesized, and tested in the same functional and binding assays as the candidate hits from V-SYNTHES (Supplementary Table S2). Out of 97 compounds from standard VLS, 16 compounds showed activity in functional assays (Extended Data Figure 7), of which 9 compounds were identified as antagonists at CB₁ with functional K_i better or equal to 10 μ M, and 5 at CB₂. Of these, 3 compounds had submicromolar antagonist K_i at CB₁, and none at CB₂. Binding affinity better than 10 μ M was detected for 8 compounds at CB₁ and 15 at CB₂ (8% and 15% hit rates respectively) (Extended Data Figure 8). Thus, hit rates for the standard VLS did not exceed 15% in any of the assays, as opposed to 33% hit rate obtained for candidate compounds selected by the V-SYNTHES approach.

Optimization of initial V-SYNTHES hits

Hits identified using V-SYNTHES have a great potential for further optimization because the combinatorial nature of the vast REAL Space of 11 Billion compounds ensures thousands of close analogs for structure-activity relationship analysis (SAR). To assess this potential, we performed the first “SAR-by-catalog” search for three of the most prominent hits (**523**, **610**, **673**) in REAL space. Chemical similarity search using ChemSpace fast algorithms selected 920 compounds within 0.3 Tanimoto distance from the hits. Hits from initial V-SYNTHES screening containing the same synthons as selected hit compounds were added to the list of similar compounds. Based on docking in the same CB₂ structural model 121 of these analogs were selected for synthesis, with 104 of selected compounds synthesized within 5 weeks (Supplementary Table S3). Testing in functional assays detected 60 analogs with potency better than 10 μ M (Extended Data Figure 9 and Supplementary Table S4) and 23 analogs with sub- μ M antagonist potency at CB₂ (13 for **523** analogs, 7 for **610**, and 3 for **673**) (Extended Data Figures 10 and 11) A series of **523** analogs yielded the most potent antagonists, with at least 5 compounds (**733**, **736**, **742**, **747**, **749**) in the low-nM range and more than 50-fold CB₂ vs. CB₁ selectivity in their binding affinity and functional potency (Figure 4). The highest, affinity was shown for compound **733** ($K_i=1.0$ nM). Like their parent V-SYNTHES hit **523**, the best analogs **733** and **747** also demonstrated high selectivity against the GPCRome-Tango panel of more than 300 receptors²³ (Extended Data Figure 12). Thus, the V-SYNTHES screen and subsequent SAR-

by-catalog allowed identification of a CB₂ selective lead series with nanomolar activity, good chemical tractability, and physical-chemical properties, without even requiring custom synthesis.

V-SYNTHES applied to ROCK1 inhibitor discovery

To assess the more broad applicability of the V-SYNTHES approach, we tested its performance on the Rho-associated coiled-coil containing protein kinase 1 (ROCK1), which is an important and challenging target in cancer drug discovery^{26,27}. We performed V-SYNTHES screen on 11B compounds with minor modifications in the selection procedure (see Methods). The benchmark comparing docking of a random compound subset of 2-component REAL space with docking of selected MEL fragments (Extended Data Figure 13) suggests enrichment EF₁₀₀~180 for ROCK1, which is comparable to EF₁₀₀~250 obtained for CB screening.

We then selected and ordered 24 fully enumerated compounds, of which 21 were synthesized and tested for functional potency and binding affinity in human ROCK1 inhibition assays (Extended Data Figure 14). Potencies better than 10 μ M were found for 6 compounds (28.5% hit rate), with 5 of these also showing binding affinities K_d < 10 μ M in the competitive binding assay. The best compound, RS-15, achieved potency IC_{50} =6.3 nM and affinity K_d =7.9 nM.

Discussion

We introduce V-SYNTHES, a new iterative approach for fast structure-based virtual screening of combinatorial compound libraries, and apply it here to discover novel antagonists chemotypes of cannabinoid CB₁ and CB₂ receptors among >11B compounds of Enamine REAL Space. In the computational benchmark, V-SYNTHES first iteration enriched the enumerated library with high-scoring candidate hits as much as 250-fold for 2-component and 460-fold for 3-component reactions, as compared to a random subset of the REAL space. Moreover, the experimental hit rate for V-SYNTHES (~33%) was twice as high compared to a full VLS of 115M diversity subset of Enamine REAL, which used ~100 times more computational resources to complete. Similarly, high hit rates and potent nanomolar antagonists were obtained by V-SYNTHES for a kinase target, ROCK1, suggesting the approach utility for different classes of protein targets.

The benefits of the V-SYNTHES modular approach in screening Giga-size libraries, while already substantial with current REAL Space, are expected to further increase in the future when the size of such libraries becomes even more prohibitive for conventional full screening. In a year, the drug-like portion Enamine REAL Space grew from ~11B to more than 19B compounds, increasing from 121 to 185 reactions and from 75,000 to 115,000 unique reactants, and will continue to grow polynomially. Thus, the library can grow as fast as a square of synthon numbers for the 2-component reactions, and even faster for 3- and more component reactions. In contrast, the V-SYNTHES computational cost increases only linearly with the number of synthons, and thus can easily accommodate the further growth of REAL Space towards Tera- and Peta-scale libraries.

Conceptually, V-SYNTHES takes advantage of the same paradigm as fragment-based ligand discovery, FBLD²⁸⁻³⁰, where binding of an anchor fragment serves as a core for growing the full drug-like compounds. Classical FBLD, however, requires experimental testing of fragment binding by highly sensitive approaches such as NMR, X-ray, or SPR, and thus is limited to smaller libraries (~1000 compounds) of smaller fragments (<200 DA). The validated fragments are then elaborated by expanding them to fill the binding pocket or connecting several fragments into one molecule, which requires elaborate custom chemistry. In contrast, V-SYNTHES avoids both the experimental testing of weakly binding fragments and custom synthesis of compounds by performing fragment enumeration in a very large but well-defined REAL chemical space, and yields drug-like compounds with affinities and potencies reliably measurable by standard biochemical assays. The apparent caveat of skipping experimental validation of initial fragments is a higher reliance on computational docking accuracy. This can, however, be compensated in several ways. First, the initial MEL compounds are small (250-350Da) and relatively rigid, which is optimal for the performance of most docking algorithms, allowing better sampling and higher success rates³¹⁻³⁴. Second, detection of strong anchor fragments and their validation in the context of full drug-like molecules makes V-SYNTHES hits highly suitable for subsequent optimization. Thus, SAR-by-catalog for several CB₂ hit analogs here yielded low-nM compounds with strong CB₂ selectivity, all achieved without requiring elaborate custom synthesis.

By design, V-SYNTHES is not limited to cannabinoid receptors (GPCRs) and ROCK1 (a kinase), but potentially can be applied to any target with a well-defined crystal or cryo-EM structure, including orphan receptors and allosteric pockets. Moreover, while this implementation uses ICM-Pro docking and employs Enamine REAL Space library, the iterative synthon-based screening algorithm can be implemented with any reliable docking-based screening platform and use any ultra-large modular library that can be represented as a combination of scaffolds and synthons. Such implementations may require custom adjustment of some parameters of the algorithm for optimal performance, opening many avenues of further exploration of this approach.

Methods

Preparation of synthons and reactions libraries

The database of reactions and corresponding synthons was kindly provided by Enamine (the version of May 2019). All reactions in the database can be separated into two categories: 2-components and 3-component reactions, based on the number of variable synthons. Synthons and reaction libraries were prepared for enumeration using ICM-Pro Molecular Modeling Software³⁵ (Molsoft LLC, San Diego). For each reaction from the reactions database a Markush structure, representing a reaction scaffold with defined attachment points for substituent synthons, was generated in a smile format. Structures of possible synthons for each R-group in each reaction were generated in 2D format with attachment points defined for enumeration. An example of 2-reagents reactions is the one-pot reductive amination of aldehydes with heteroaromatic amines³⁶, as shown in Extended Data Figure 15a. An example of 3-reagent reaction is one-pot formation of thiazoles via asymmetrical thioureas³⁷, shown in Extended Data Figure 15b.

Enumeration of Combinatorial Library

Enumeration of combinatorial libraries was performed using combinatorial chemistry tools implemented in ICM-Pro³⁵. Markush structures for enumeration were derived from reaction SMARTS provided by Enamine.

Generation of Minimal Enumeration Library (MEL)

Minimal Enumeration Library was generated to generate all possible synthon-scaffold combinations in Enamine REAL Space. Each compound in the MEL library comprises a reaction scaffold enumerated with a single synthon, while other attachment points are replaced with the minimal synthons, or “caps”. Minimal chemically feasible synthons for every substituent in each reaction were selected as either methyl or phenyl, later one in case the reaction required an aromatic group. Minimal synthon atoms were labeled as ¹³C isotopes to facilitate computational analysis of docking poses (Extended Data Figure 2).

In 2-component Minimal Enumeration Library generation, filters on molecular weight and cLogP were applied to remove MEL compounds with MW>400 and cLogP >5, which would likely result in fully enumerated compounds that violate Lipinski's rule of 5. For 3-component reactions, the size filters were set to MW<350 on the first iteration of V-SYNTHES and to MW<425 on the second.

Generation of random enumerated library

To generate random subsets of the REAL database for internal benchmarking was performed by enumeration of randomly selected synthons from each reaction. To create the 1 Million compound library for 2-component reactions, 1% of synthons (total of 6418 synthons) were randomly selected, which represented each R group in each reaction. For 3-component reactions, 0.47% of synthons (total of 512 synthons) were randomly selected for the 500K library, with no less than 1 synthon per Markush R group. The random libraries were filtered by Lipinski's rules of five.

Selection of MEL candidates for CB₁/CB₂ for full enumeration

To select MEL candidates for further enumeration, the score and docking pose of each MEL candidate were analyzed. The fragments were ranked by score and the top 1% were kept for further investigation. To detect “productive” vs. “non-productive” compound poses, the algorithm calculates the distances between the cap atoms of docked MEL candidates and the selected atoms (or dummy atoms) marking the dead-end sub-pocket in the protein binding site. For the CB₂ receptor pocket, three dead-end points were used to define potentially “non-productive” MEL ligands: water molecule from the crystal structure and two dummy atoms, one placed between residues F106 and K109, another between residues H95 and L182. MEL compounds for which their cap atoms closer than 4 Å to the “dead-end” points were excluded from further consideration. Furthermore, to ensure the diversity of the final library, the best MEL candidates were filtered in a way that the final selection did not contain more than 20% of the MEL candidates from the same reaction.

For 2-component reactions, 819 best MEL candidates were selected for further enumeration resulting in 1M library of full compounds. For 3-component reactions, two rounds of

enumerations were required to arrive at full molecules. In the first round, 1043 best MEL candidates were used to produce 500K molecules with two real synthons and one minimal cap. After docking and analysis of these ligands, 4739 best molecules were selected for the final enumeration step resulting in 500K fully enumerated molecules.

Receptor model preparation for CB₂

Both V-SYNTHES and standard VLS employed a structural model based on CB₂R crystal structure with an antagonist AM10257 at 2.8 Å resolution (PDB ID 5ZTY)²². The structure was converted from PDB coordinates to the internal coordinates object using the ICM-Pro conversion tool by restoring missing heavy atoms and hydrogens, locally minimizing polar hydrogens, and optimizing His, Asn, and Gln side chains protonation state and rotamers. In the final step of selection, we also used ligand-optimized structural models for redocking of top 1% hits. These refined models were generated in a ligand-guided receptor optimization procedure (LiBERO)³⁸, which refined the sidechains and water molecules within the 8 Å radius from the orthosteric binding pocket. Two binding modes for CB₂ receptor binding pocket were prepared: one guided by 20 known antagonists and another by 20 agonists, selected from ChEMBL high-affinity ligands for CB₂ (ChEMBL253, pK > 8). These compounds, along with 200 decoy molecules selected from CB₂ receptor decoy database (GDD)³⁹ were docked into the refined conformers. The conformers yielding the best AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curves were selected as the best LiBERO models. The two LiBERO models, along with the crystal structure model, were combined into one 4D model as described previously⁴⁰. The 4D model was used for screening in both V-SYNTHES iterative algorithm and standard VLS. Unlike V-SYNTHES, standard VLS used a preassembled library of 115 Million of REAL compounds, including 100M of a lead-like subset of REAL and a diversity REAL subset of 15M drug-like compounds⁴¹.

Docking and VLS for CB₂

Docking simulations in both V-SYNTHES and standard VLS were performed using ICM-Pro molecular modeling software (Molsoft LLC)³⁵. Docking involves an exhaustive sampling of the molecule conformational space in the rectangular box that comprised the CB₂ orthosteric binding pocket and was done with the thoroughness parameter set to 2. Docking uses biased probability Monte Carlo (BPMC) optimization of the compound's internal coordinates in the pre-calculated grid energy potentials of the receptor. The 4D model of the receptor pocket described above was used to sample 3 slightly different receptor conformations in a single docking run as implemented in ICM-Pro (Molsoft LLC). Before the final selection of hits for experimental testing, the top 30K compounds from the screen were re-docked into the model with higher thoroughness (5) to assure their comprehensive sampling.

V-SYNTHES enrichment factor for CB₂

To evaluate the efficiency of the V-SYNTHES approach and compare it with standard VLS, we introduced an “enrichment factor” that provides a quantitative measurement of how the final library on Step 4 of the algorithm is enriched in hits as compared to a library of the same size generated as a random subset of the Enamine REAL space. For

2-component reactions (500 M compounds), we compared random and enriched libraries of 1M compounds. For 3-component reactions (total 10.5B compounds), we compared random and enriched libraries of 0.5M compounds. The enrichment is calculated for hits with docking scores equal to or better than a certain threshold X, and is defined as the following ratio:

$$\text{Enrichment factor}(X) = \frac{N \text{ of hits with scores } < X \text{ in SYNTHESES}}{N \text{ of hits with scores } < X \text{ in standard VLS}}$$

Enrichment factor at the docking score threshold that selects 100 candidate hits, designated EF100, can be used as a single value practical metric of the algorithm performance.

Generating initial SAR for selected CB₂ hits

Chemical search for analogs of best compounds **523**, **610**, and **673** in REAL Space was performed using REALSpaceNavigator¹⁶. Compounds with Tanimoto distance less than 0.3 (< 0.4 for BRI-15673) were selected for docking. Following criteria were used to select top-scoring compounds for each parent molecule: docking scores better than -30 (-25 for **673**), cLogP < 5, cLogS > -5, MW < 500, and Tanimoto distance to known CB₁/CB₂ ligands > 0.3. In addition, the 20K top hits from initial V-SYNTHESES screening were re-analyzed and the best molecules generated from the same fragments as **523**, **610**, and **673** were added to the final list. The number of analogs selected for synthesis: 49 for **523** (49 synthesized), 42 for **610** (38 synthesized), and 30 compounds for **673** (17 compounds synthesized).

Parallel Synthesis

Parallel one-pot synthesis for all compounds in this study was performed by Enamine in 5 weeks with >90% purity guaranteed as described in Supplementary Information Methods. This includes (1) candidate CB compounds from the initial V-SYNTHESES round (60 synthesized out of 80 ordered), (2) SAR-by-catalog compounds (104 out of 121), (3) compounds from the benchmark full screen of 115 REAL diversity library (97 out of 109), and (4) ROCK1 candidate compounds (21 synthesized out of 24 ordered):

Functional potency in CB₁/CB₂ Tango assays

The Tango arrestin recruitment assays were performed as previously described²³. Briefly, HTLA cells were transiently transfected with human CB₁ or CB₂ Tango DNA construct overnight in DMEM supplemented with 10 % FBS, 100 µg/ml streptomycin and 100 U/ml penicillin. The transfected cells were then plated into Poly-L-Lysine coated 384-well white clear bottom cell culture plates in DMEM containing 1% dialyzed FBS at a density of 10,000-15,000 cells/well. After 6 hours of incubation, the plates were added with drug solutions prepared in DMEM containing 1% dialyzed FBS for overnight incubation. Specifically for the antagonist assay, 100 nM of CP55940 was added after 30 minutes of incubation of the drugs. On the day of assay, medium and drug solutions were removed and 20 µL/well of BrightGlo reagent (Promega) was added. The plates were further incubated for 20 min at room temperature and counted using a Wallac TriLux Microbeta counter (PerkinElmer). Results were analyzed using GraphPad Prism 9. Each experiment was

performed in triplicate and functional K_i values were determined from three independent experiments and are expressed as the mean of the three values.

Radioligand binding in CB₁/CB₂ binding assays

The affinities (K_i) of the new compounds for rat CB₁ receptor as well as for human CB₂ receptors were obtained by using membrane preparations from rat brain or HEK293 cells expressing hCB₂ receptors, respectively, and [³H]CP-55,940 as the radioligand, as previously described.^{42,43} Results from the competition assays were analyzed using nonlinear regression to determine the IC₅₀ values for the ligand; K_i values were calculated from the IC₅₀ using GraphPad Prism 9. Each experiment was performed in triplicate and K_i values were determined from three independent experiments and are expressed as the mean of the three values.

PRESTO-Tango GPCRome

Screening of the compounds in the PRESTO-Tango GPCRome was performed as previously described²³ with modifications. First, HTLA cells were plated in poly-L-lysine coated 384-well white plates in DMEM containing 1% dialyzed FBS for 6 hours. Next, the cells were transfected with 20 ng/well PRESTO-Tango receptor DNAs overnight. Then, the cells were added with 10 μ M drugs without changing the medium and incubated for another 24 hours. Each target was designed to have 4 wells for basal and 4 wells for sample. The remaining steps of the PRESTO-Tango protocol²³ were followed. The results were plotted as fold of average basal against individual receptors in the GraphPad 9.0 software. For the receptors that had > 3-fold of basal signaling activity, assays were repeated as a full dose-response assay and the results were plotted as a percentage of reference compounds.

V-SYNTHES applied to ROCK1 screen

MEL library was docked into the ROCK1 crystal structure (PDB:2etr)⁴⁴ prepared in ICM-Pro. The 20K best-scoring fragments were then screened for their hydrogen bond interactions with the hinge region of ROCK1, residues E154 and M156. To eliminate potentially “non-productive” fragments in the enumeration step, all fragments with capping atoms within 4.6Å distance from these hinge region residues were removed, leaving about 5K compounds for enumeration with full synthons. Docking of the 1 Million fully enumerated compounds resulted in the top 30k compounds with docking scores ranging between -35 to -50 kJ/mol. The vast majority of them (>99%) retained hydrogen bonding to hinge region residues, showing that the full molecules maintain the binding properties predicted for MEL fragment selection. The remaining compounds were filtered with PAINS score, drug-likeness properties, chemical diversity as well as ligand interaction diversity to sample different binding modes in the pocket. 24 compounds were selected for purchase from Enamine, of which 21 were successfully synthesized with a purity >90% and were delivered in under 6 weeks.

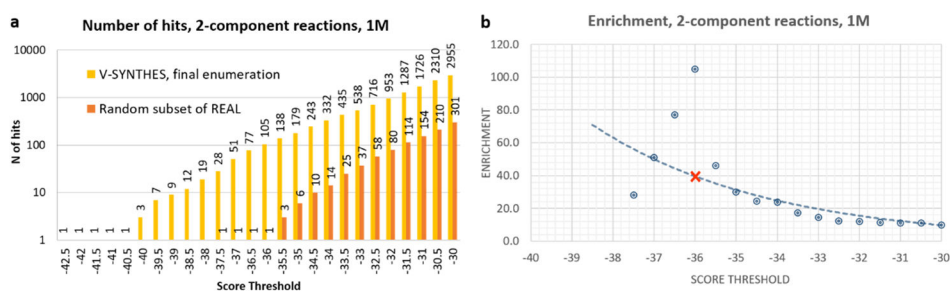
ROCK1 functional and binding assays

The HotSpot radiometric assay (Reaction Biology Corporation) measures inhibition of ROCK1 catalytic activity toward specific peptide substrate

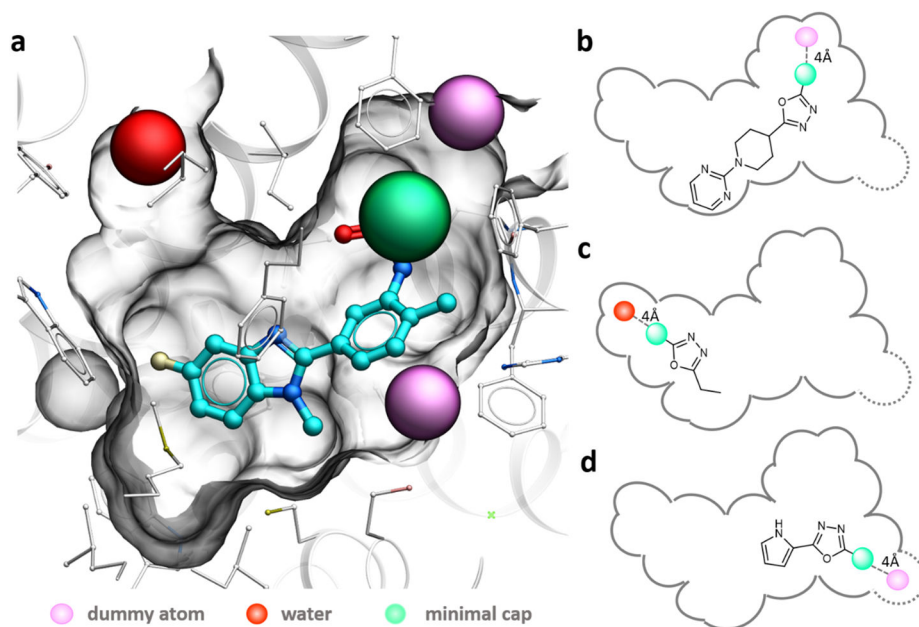
[KEAKEKRQEQIAKRRRLSSLRASTSKSGGSQK], which is monitored by P81 filter-binding methods⁴⁵. All compounds were tested in triplicate at starting concentration of either 100 μ M or 90 μ M in the presence of 1 μ M ATP and diluted three-fold for a total of 10 doses.

The KdElect assay (Eurofins/DiscoverX) measures quantitative binding (Kd) of compounds to ROCK1 in competition with an immobilized active-site directed ligand. Binding is determined by measuring the amount of kinase captured by immobilized ligands versus the control samples through the use of qPCR. Soluble compounds specifically binding to ROCK1 prevent the immobilized ligand from binding. Our compounds were tested in triplicate in an 11-dose response curve at a starting concentration of 30 μ M. IC₅₀ was calculated and graphed using a nonlinear regression curve in Graphpad Prism 8.

Extended Data

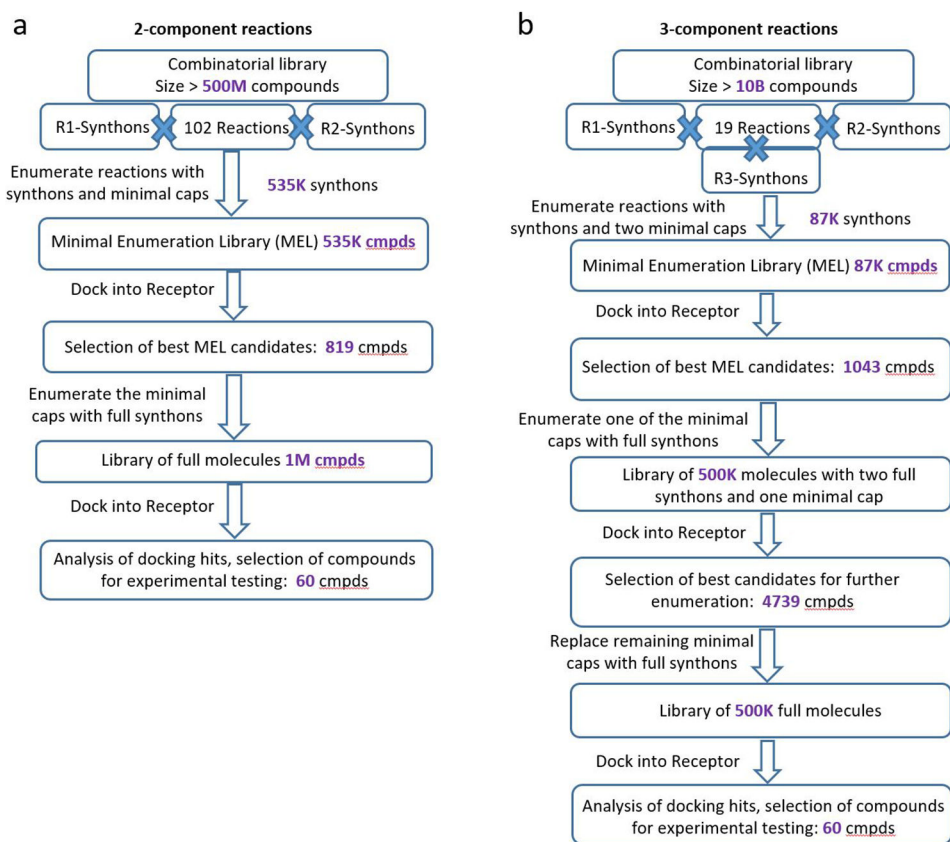


Extended Data Figure 1. Evaluation of SYNTHES performance on CB₂ receptor with only docking score (without considering docking pose of MEL candidates in the binding pocket). (a) The number of hits at each score threshold from V-SYNTHES and standard VLS (b) Enrichment in V-SYNTHES vs. Standard VLS at different score thresholds, with the red x-mark showing threshold that yields 100 V-SYNTHES hits in the 2-component library.



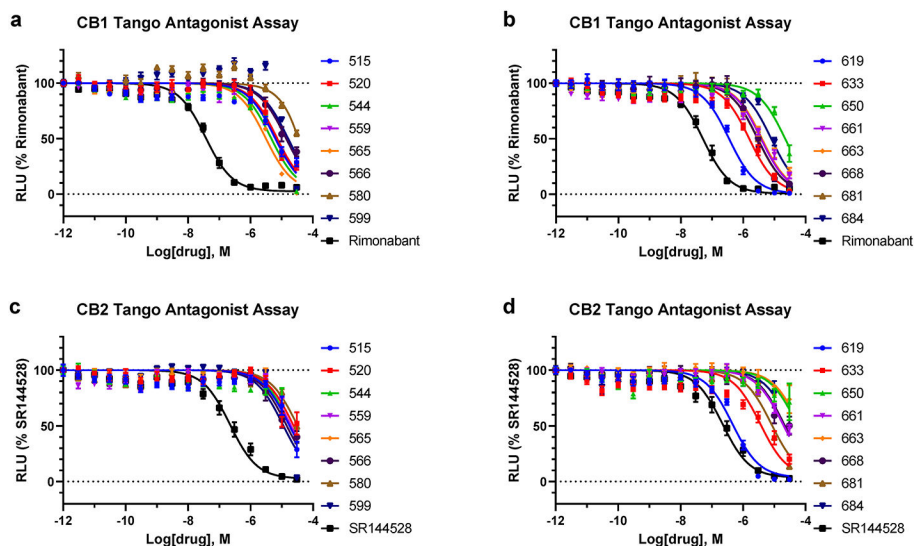
Extended Data Figure 2. Binding pocket of CB₂ with selected dead-end atoms.

a) 3D illustration of a MEL compound binding pose (carbon atoms colored cyan) with a “non-productive” pose. **(b,c,d)** 2D schematics showing other possible non-productive cases, including dead-end sub-pockets. Dead-end water-colored red, pseudoatoms colored magenta.



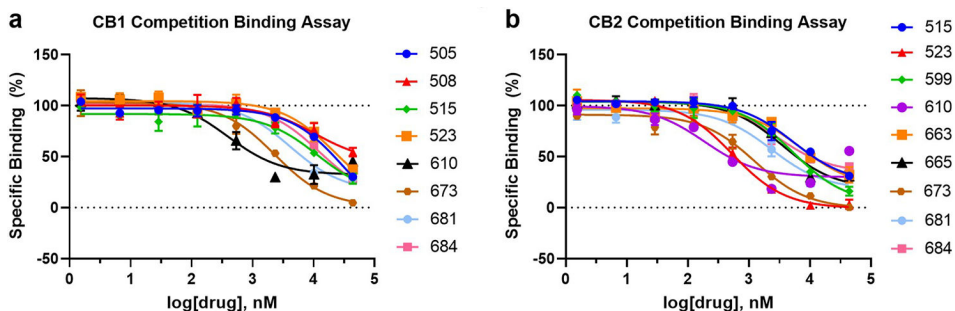
Extended Data Figure 3. Details of practical application V-SYNTHES algorithms to CB receptors screening

in (a) 2-component and (b) 3-component reaction cases.



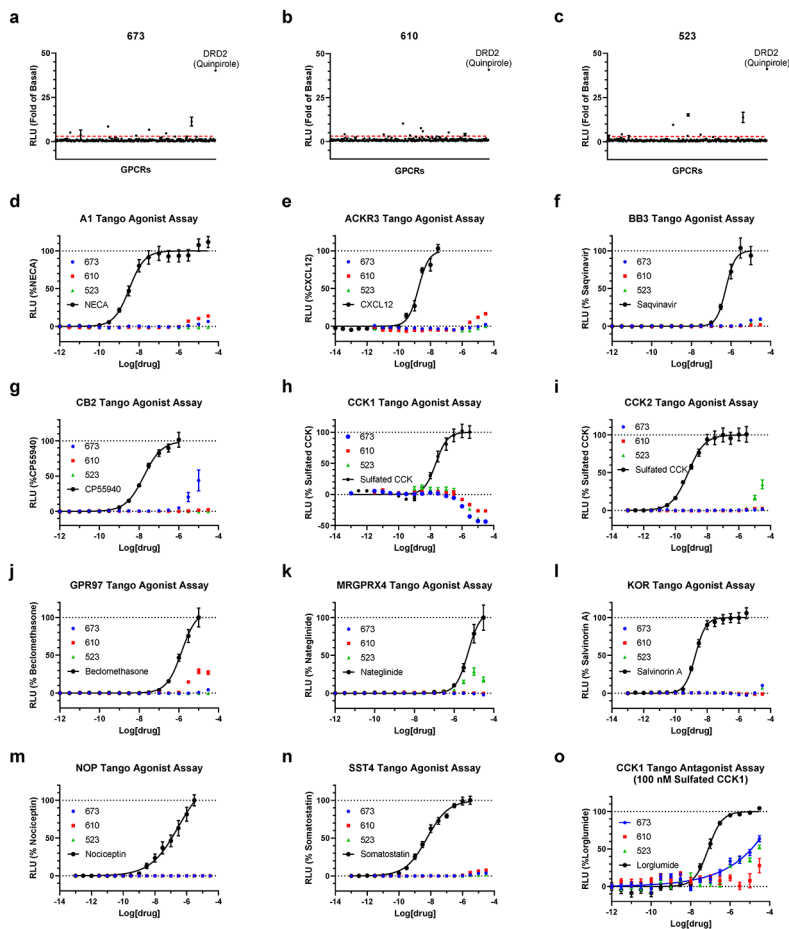
Extended Data Figure 4. Concentration-response curves for V-SYNTHES hits in functional assays at CB₁ and CB₂ receptors (except those shown in main text Figure 3).

β -arrestin recruitment Tango assays were performed to assess antagonist activity of the compounds in (a-b) CB₁ and (c-d) CB₂ receptors. The compounds rimonabant or SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the dual CB₁/CB₂ CP55,940 agonist. The data points are presented as mean \pm SEM with n=3 independent experiments, each one carried out in triplicate.

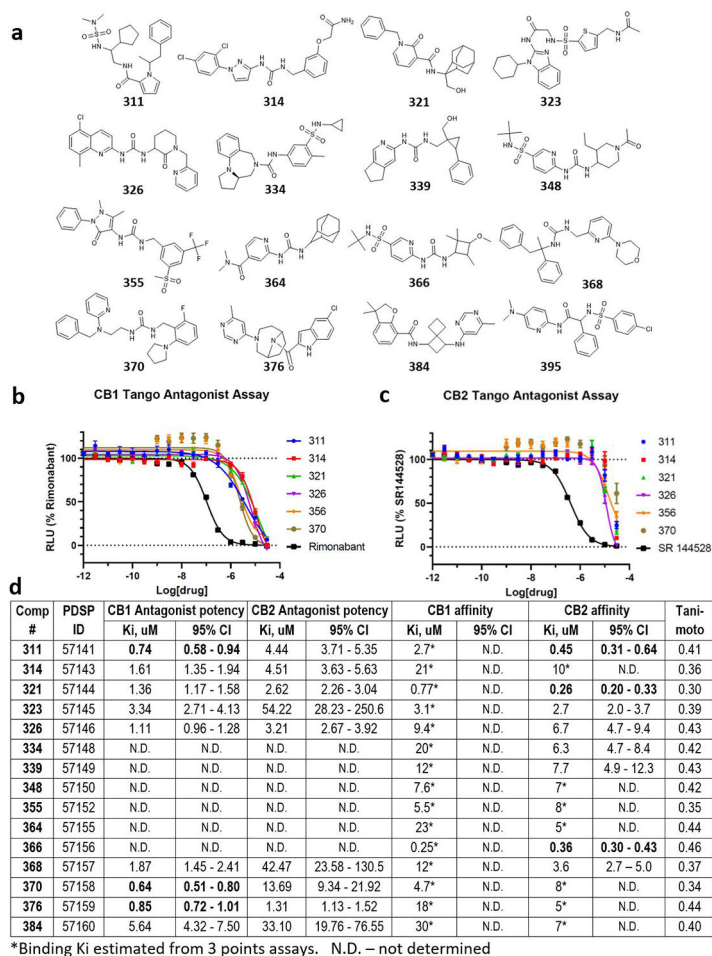


Extended Data Figure 5. Competition binding curves for the best CB₂ hit compounds from V-SYNTHES.

Radioligand binding assays were used to assess the binding affinities in rCB1 (a) and hCB₂ (b). [³H]CP-55,940 was used as the radioligand. The data were presented as mean \pm SEM with n=3 independent experiments, each one carried out in triplicate.

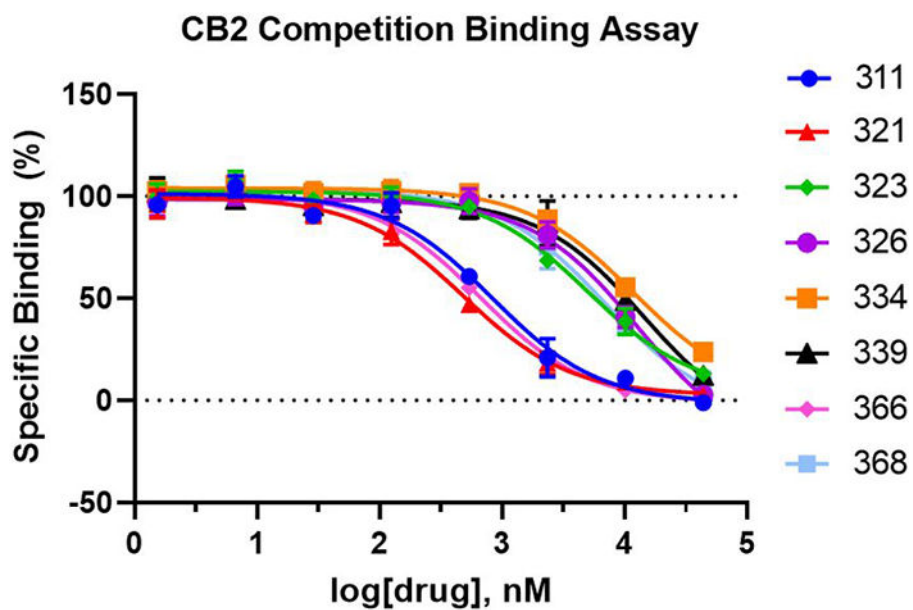


Extended Data Figure 6. Assessment of off-target selectivity for the best V-SYNTHES CB₂ hits. (a-c) Screening of compounds 673, 610 and 523 at 10 μM concentrations in GPCRome-Tango assays for >300 receptors. Dopamine D₂ (DRD2) and 100 nM Quinpirole served as an assay control. The data are presented as mean ± SEM (n=4) and the values of fold of basal > 3 are marked as significant hits. (d-o) Follow-up dose-response curves for targets with >3 fold increased activity. Known agonists or antagonist that showed activity served as positive controls. The data points are presented as mean ± SEM with n=3 independent experiments, each assay carried out in triplicate.



Extended Data Figure 7. Identification and characterization of CB1 and CB2 hits from standard VLS of 115M Enamine REAL compounds.

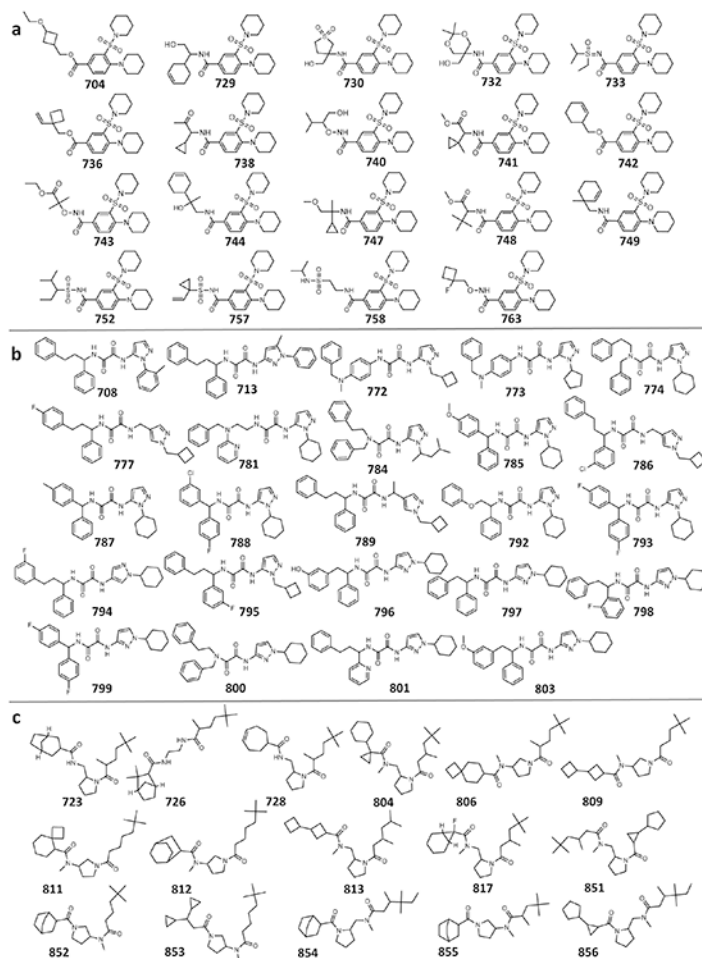
(a) Chemical structures of the hits from the standard VLS. (b-c). Concentration-response curves of the best hits in β -arrestin recruitment Tango assays for antagonist activity at CB₁ (b) and CB₂ (c) receptors. The compounds rimonabant or SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the dual CB₁/CB₂ CP55,940 agonist. The data points are presented as mean \pm SEM with n=3 independent experiments, each one carried out in triplicate. (d) Functional potencies and binding affinities of the hit compounds from standard VLS. The 95% Confidence Intervals (CI) were calculated from n=3 independent assays, with 16 dose-response points for functional K_i values and 8 dose-response points for affinity K_i values, except for values marked with *, roughly estimated from 3-point assays.



Extended Data Figure 8. Competition binding curves for the best CB₂ hit compounds from standard VLS.

Radioligand binding assays were used to assess the binding affinities in hCB₂.

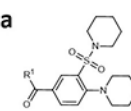
[³H]CP-55,940 was used as the radioligand. The data were presented as mean ± SEM with n=3 independent experiments, each one carried out in triplicate.















Extended Data Figure 9. Chemical structures for series of the SAR-by-catalog analogs of antagonists, discovered by V-SYNTHES.

Shown are 60 analogs of **523** (a), **610** (b), and **673** (c) with inhibitory activity >40% in the single point functional assays. All 104 analogs tested are shown in Supplementary Information Table S3.

a

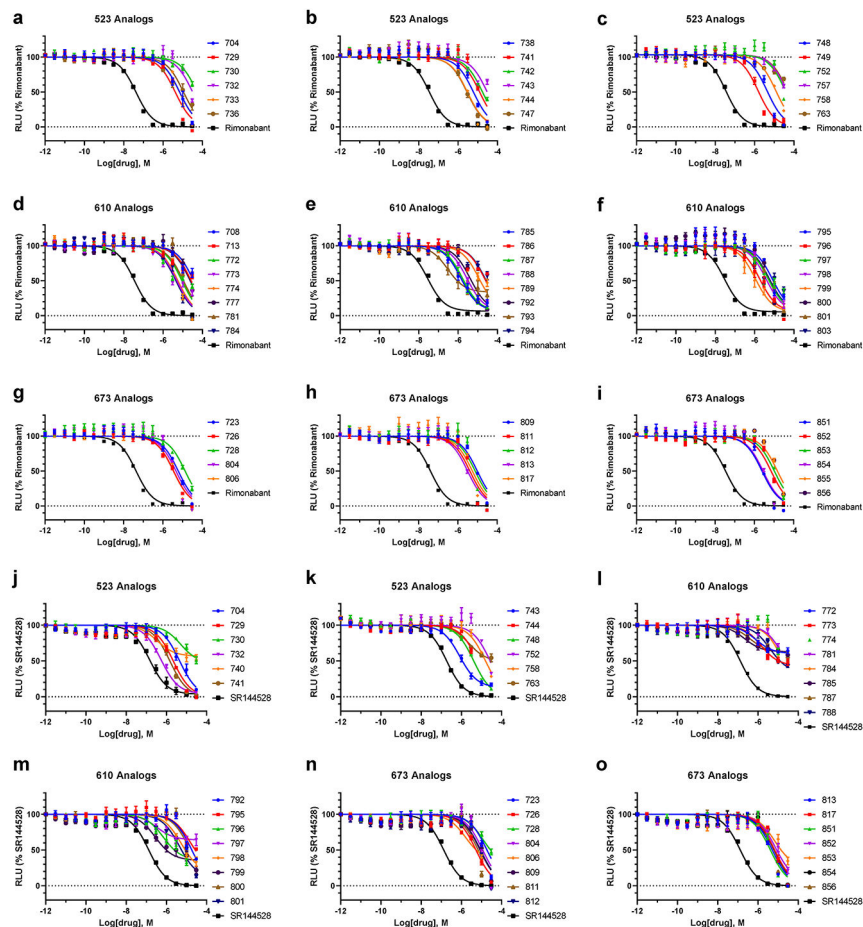


b

Cmpd #	PDSP ID	R ₁	CB ₁ Antagonist potency		CB ₂ Antagonist potency		CB ₁ affinity		CB ₂ affinity		Tanimoto	
			K _i , nM	95% CI	K _i , nM	95% CI	K _i , nM	95% CI	K _i , nM	95% CI	parent	CB _{1/2} ligands
729	59026		542	454-647	388	313-481	N.D.		N.D.		0.25	0.33
732	59029		3994	3215-5000	138	112-171	1500	969-2323	10.9	3.8-30.9	0.30	0.37
733	59030		871	720-1051	10.9	9.26-12.89	43.2	28.2-66.1	1.0	0.3-3.9	0.26	0.35
736	59033		1185	868-1603	48.5	38.6-61.0	140	105-186	3.6	1.7-7.7	0.25	0.37
738	59035		856	725-1009	125	105-148	23.1	13.9-38.6	2.1	1.5-2.9	0.25	0.30
740	59037		16490	11830-24970	246	132-452	1541	923-2573	74.4	43.1-128	0.17	0.37
741	59038		1781	1470-2157	293	241-356	N.D.		N.D.		0.30	0.36
742	59039		2340	1878-2919	120	101-144	394	281-551	7.3	5.2-10.3	0.29	0.37
743	59040		607	257-1510	203	155-264	215	144-322	30.6	20.5-45.8	0.25	0.41
747	59044		455	373-558	9.6	8.58-10.8	228	172-303	1.3	0.5-3.6	0.28	0.33
749	59046		209	177-248	49.2	42.1-57.6	689	472-1004	8.5	3.3- 22	0.21	0.30
763	59060		322	145-747	441	262-748	N.D.		N.D.		0.17	0.37
SR144528					40.1	38.2 - 42.1						
AM10257							10.3	6.1-17.5	1.16	0.88-1.5		

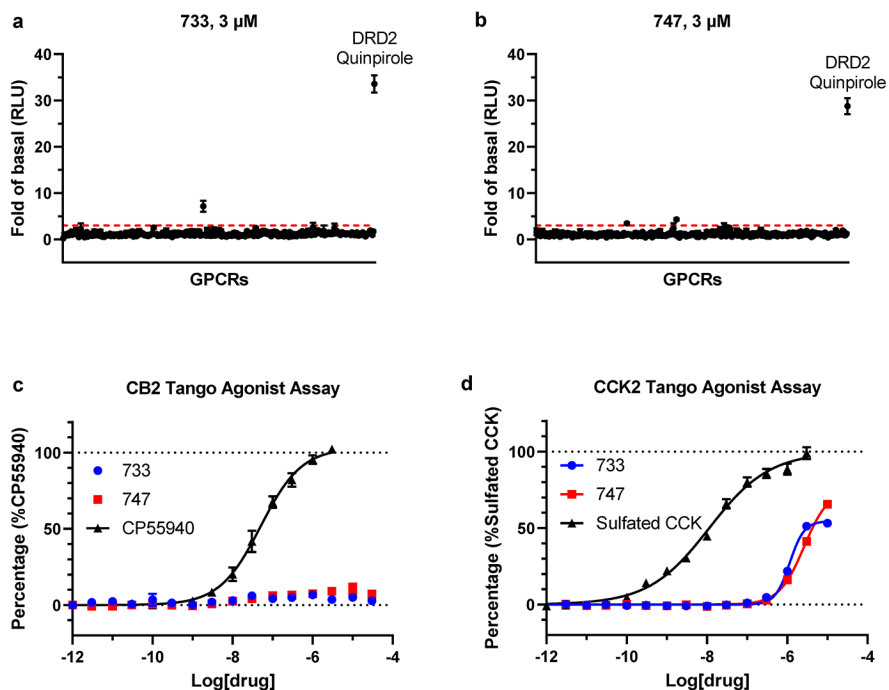
Extended Data Figure 10. Functional potency and binding affinity assessment of the SAR-by-catalog analogs of the antagonist 523, discovered by V-SYNTHESIS.

Table compounds with CB₂ potency better than 500 nM are shown, antagonists with affinities better than 10 nM highlighted in bold, >50-fold selective by italic. Functional K_i values and 95% Confidence Intervals were calculated from n=4 independent assays with 16 dose-response points. Affinity K_i values and 95% Confidence Intervals were calculated from n=3 independent assays with 8 dose-response points.



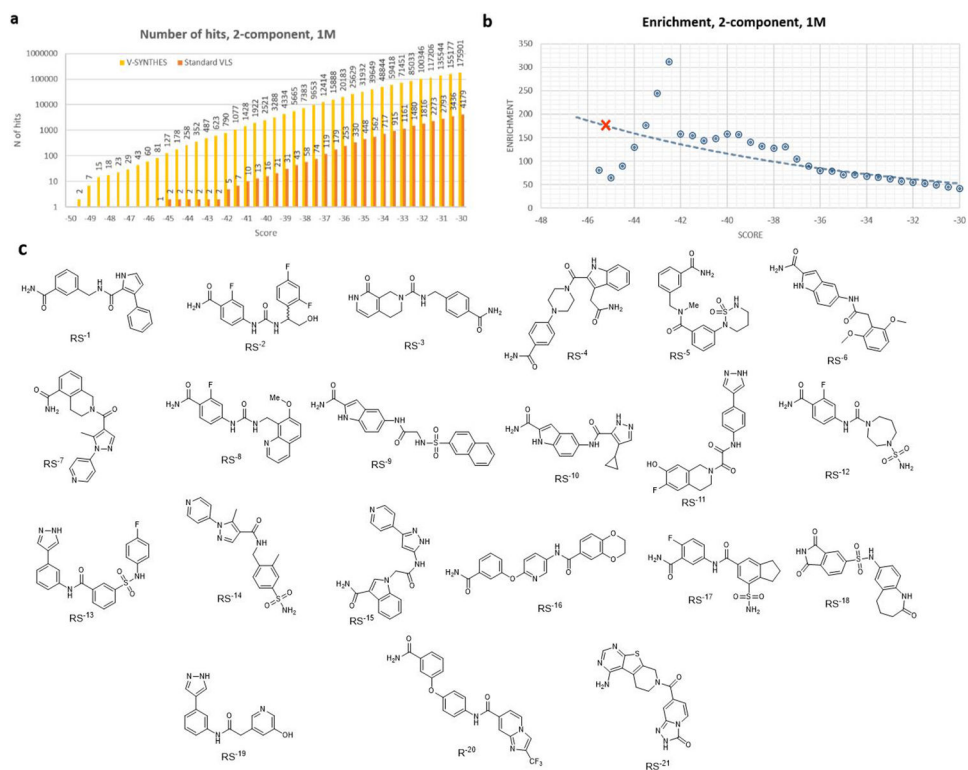
Extended Data Figure 11. Concentration-response curves for series of the SAR-by-catalog analogs of 523, 610 and 673 antagonists, discovered by V-SYNTHES.

The β -arrestin recruitment Tango assays were performed to assess the antagonist activity of the best hits at CB₁ (a-i), and CB₂ (j-o) receptors. Note that the six best analogs of **523** shown in Figure 4 are excluded here. The compounds rimonabant and SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the CP55,940 agonist. The data were presented as mean \pm SEM with n=3 independent experiments, each run carried out in triplicate.

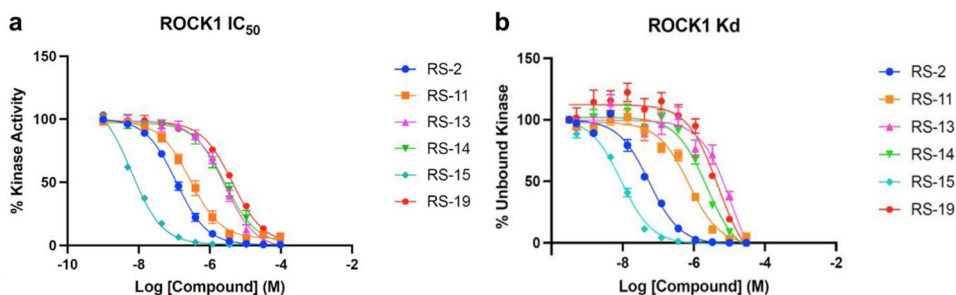


Extended Data Figure 12. Assessment of off-target selectivity for the best SAR-by-catalog compounds 733 and 747.

(a-b) Screening of compounds 733 and 747 in GPCRome-Tango assay for >300 receptors at 10 μM concentrations. Dopamine D₂ (DRD2) and 100 nM Quinpirole served as an assay control. The data are presented as mean ± SEM (n=4) and the values of fold of basal > 3 marked as significant hits. (c-d) Follow-up dose-response curves for targets with >3 fold increased activity. Known agonists that showed activity served as positive controls. The data were presented as mean ± SEM with n=3 independent experiments, each run carried out in triplicate.



Extended Data Figure 13. Application of V-SYNTHES to the discovery of ROCK1 inhibitors. (a-b) Computational assessment of V-SYNTHES performance vs standard VLS. (a) The number of candidate hits at each score threshold from V-SYNTHES and standard VLS. (b) Enrichment in V-SYNTHES vs. standard VLS at different score thresholds, with the red x-mark showing threshold that yields 100 hits in the 2-component library. (c) Chemical structures of all selected by V-SYNTHES and synthesized compounds for ROCK1 kinase.

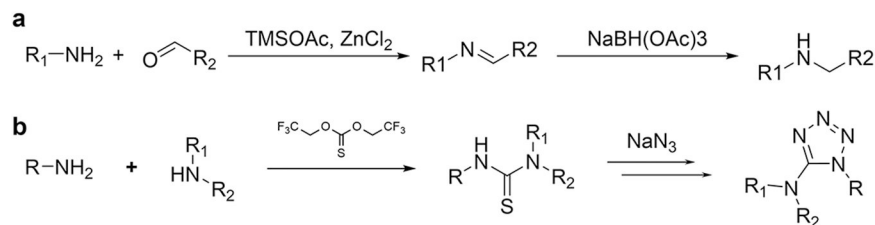


c

Compound ID	ROCK1 IC ₅₀ (μM)	95% CI, (μM)	ROCK1 K _d (μM)	95% CI, (μM)
RS-1	>100	N/D	N/D	N/D
RS-2	0.11	0.10 - 0.13	0.055	0.046 - 0.067
RS-3	>100	N/D	N/D	N/D
RS-4	>100	N/D	N/D	N/D
RS-5	~14*	N/D	N/D	N/D
RS-6	>100	N/D	N/D	N/D
RS-7	>100	N/D	N/D	N/D
RS-8	~17*	N/D	N/D	N/D
RS-9	>100	N/D	N/D	N/D
RS-10	~26*	N/D	N/D	N/D
RS-11	0.28	0.22 - 0.35	1.83	0.66 - 4.73
RS-12	~16*	N/D	N/D	N/D
RS-13	2.59	2.01 - 3.32	11.7	6.06 - 25.2
RS-14	2.74	2.07 - 3.62	2.31	1.80 - 2.98
RS-15	0.0063	0.0055 - 0.0072	0.0079	0.0077 - 0.011
RS-16	~16*	N/D	N/D	N/D
RS-17	>100	N/D	N/D	N/D
RS-18	>100	N/D	N/D	N/D
RS-19	4.35	3.47 - 5.44	4.88	3.04 - 7.99
RS-20	>100	N/D	N/D	N/D
RS-21	>100	N/D	N/D	N/D

Extended Data Figure 14. Experimental characterization of candidate ROCK1 inhibitors predicted by V-SYNTHES.

Full dose-response curves for the ROCK1 hits in (a) functional potency and (b) binding affinity at human ROCK1. The data points are presented as mean ± SEM from n=3 independent experiments, each run carried out in triplicate. (c) Values of binding affinities and functional potencies for all candidate compounds predicted by V-SYNTHES. Bold font highlight hits with IC₈₀<10 μM. Estimated values for curves that did not allow accurate fitting are marked with *.



Extended Data Figure 15. Examples of typical Enamine REAL reactions.

(a) 2-component reaction (b) 3-component reaction

Extended Data Table 1.
Potencies and affinities of V-SYNTHES hits in functional and binding assays at CB₁ and CB₂ receptors.

Sub-micromolar hits are shown in bold, selective by italic. The 95% Confidence Intervals (CI) were calculated from n=3 independent assays, with 16 dose-response points for functional K_i values and 8 dose-response points for affinity Ki values, except for values marked with *, roughly estimated from 3-point assays. Potencies are measured in assays running in antagonist mode, except for those marked & that were measured in agonist mode. N/D stands for Not Determined.

#	PDSP ID	CB ₁ Antagonist potency		CB ₂ Antagonist potency		CB ₁ affinity		CB ₂ affinity		Tanimoto distance
		K _i , uM	95% CI	K _i , uM	95% CI	K _i , uM	95% CI	K _i , uM	95% CI	
505	56707	0.28	0.22 - 0.36	0.54	0.43 - 0.67	16.4	8.6 - 31.3	1*	N/D	0.38
515	56731	0.94	0.76 - 1.16	3.81	2.89 - 5.09	6.1	2.9 - 13.0	2.85	1.9 - 4.1	0.39
520	56717	1.07	0.84 - 1.37	5.20	3.82 - 7.22	11.6	3.7 - 35.7	12.8	4.8 - 34.2	0.40
523	56737	1.82	1.46 - 2.28	1.59	1.27 - 1.98	12.0	5.4 - 26.7	0.85	0.69 - 1.05	0.39
544	56724	0.69	0.57 - 0.84	7.78	4.66 - 16.8	5.0 - 7.2*	N/D	2.5*	N/D	0.34
559	56715	0.98	0.80 - 1.20	4.25	3.15 - 5.90	N/D*	N/D	12.2	2.1-69.6	0.43
565	56684	0.46	0.40 - 0.54	3.77	2.71 - 5.53	4.5*	N/D	13.6	8.4-22.0	0.37
566	56708	2.05	1.63 - 2.60	4.04	3.02 - 5.48	6.9*	N/D	1.2	0.84 - 1.57	0.43
580	56727	5.80	4.55 - 7.55	6.92	5.51 - 8.80	1.0-9.0*	N/D	1.5*	N/D	0.36
599	56723	2.33	1.82 - 3.01	2.44	2.06 - 2.89	26.5*	N/D	10.4	7.1 - 15.1	0.34
610	56696	0.76	0.62 - 0.93	4.17	3.14 - 5.62	0.62	0.34 - 1.13	0.28	0.12 - 0.69	0.31
619	56695	0.05	0.04 - 0.06	0.11	0.09 - 0.13	45*	N/D	0.9-2.5*	N/D	0.42
633	56726	0.23	0.19 - 0.28	1.53	1.18 - 1.98	10*	N/D	0.7-0.9*	N/D	0.50

#	PDSP ID	CB ₁ Antagonist potency		CB ₂ Antagonist potency		CB ₁ affinity		CB ₂ affinity		Tanimoto distance
		K _i , uM	95% CI	K _i , uM	95% CI	K _i , uM	95% CI	K _i , uM	95% CI	
650	56725	3.22	2.61 - 4.01	12.2	7.85 - 20.7	45*	N/D	0.9-2.5*	N/D	0.48
661	56685	0.55	0.43 - 0.70	4.37	3.37 - 5.74	19*	N/D	4.0	2.4- 6.7	0.39
663	56687	0.59	0.46 - 0.75	14.5	9.89 - 23.0	12.5*	N/D	14.3	6.6 - 30.9	0.36
665	56732	0.39	0.32 - 0.47	0.82	0.71 - 0.95	>7*	N/D	6.7	4.2 - 10.6	0.47
668	56691	0.43	0.33 - 0.56	4.78	3.60 - 6.42	5.5-6.9*	N/D	5.2	2.5 - 11.0	0.40
673	56683	0.97	0.84 - 1.14	3.66	2.98 - 4.51	4.2	2.9 - 6.0	2.2	1.4 - 3.4	0.46
681	56701	0.42	0.32 - 0.55	1.86	1.52 - 2.30	8.2	5.2 - 12.7	4.2	2.5 - 7.2	0.42
684	56689	1.16	0.93 - 1.43	7.28	4.50 - 14.4	25.5	16.6 - 39.0	5.3	3.5 - 8.1	0.48
SR144528		N/D	N/D	0.052	0.041 - 0.066					
Rimonabant		0.006	0.005 - 0.008	N/D	N/D					
CP55940 ^{&}		0.017		0.028						

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

The study was funded by National Institute on Drug Abuse grants R01DA041435 and R01DA045020 (V.K. and A.M.), National Institute of Mental Health Grant R01MH112205 and Psychoactive Drug Screening Program (B.L.R), and the Michael Hooker Distinguished Professorship (B.L.R). B.H. was supported by NIGMS T32-GM118289. We would like to thank the USC Center for Advanced Research Computing, and the Google Cloud Platform for Higher Education and Research for providing computational resources.

References

1. Shoichet BK & Kobilka BK Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol Sci* 33, 268–72 (2012). [PubMed: 22503476]
2. Katritch V, Cherezov V & Stevens RC Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol* 53, 531–56 (2013). [PubMed: 23140243]
3. Renaud J-P et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery* 17, 471–492 (2018). [PubMed: 29880918]

4. Congreve M, de Graaf C, Swain NA & Tate CG Impact of GPCR Structures on Drug Discovery. *Cell* 181, 81–91 (2020). [PubMed: 32243800]
5. Stein RM et al. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* 579, 609–614 (2020). [PubMed: 32040955]
6. Lyu J et al. Ultra-large library docking for discovering new chemotypes. *Nature* 566, 224–229 (2019). [PubMed: 30728502]
7. Grygorenko OO et al. Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. *iScience* 23, 101681 (2020). [PubMed: 33145486]
8. Gorgulla C et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 580, 663–668 (2020). [PubMed: 32152607]
9. Graff DE, Shakhnovich EI & Coley CW Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem Sci* 12, 7866–7881 (2021). [PubMed: 34168840]
10. Engels MF & Venkatarangan P Smart screening: approaches to efficient HTS. *Current opinion in drug discovery & development* 4, 275–283 (2001). [PubMed: 11560058]
11. Villoutreix BO, Eudes R & Miteva MA Structure-based virtual ligand screening: recent success stories. *Comb Chem High Throughput Screen* 12, 1000–16 (2009). [PubMed: 20025565]
12. Abagyan R & Totrov M High-throughput docking for lead generation. *Curr Opin Chem Biol* 5, 375–82 (2001). [PubMed: 11470599]
13. Irwin JJ & Shoichet BK Docking Screens for Novel Ligands Conferring New Biology. *J Med Chem* 59, 4103–20 (2016). [PubMed: 26913380]
14. Ertl P Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J Chem Inf Comput Sci* 43, 374–80 (2003). [PubMed: 12653499]
15. Bohacek RS, McMartin C & Guida WC The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev* 16, 3–50 (1996). [PubMed: 8788213]
16. Enamine. <https://enamine.net/library-synthesis/real-compounds/real-space-navigator>. (2020).
17. Guzmán M Cannabinoids: potential anticancer agents. *Nat Rev Cancer* 3, 745–55 (2003). [PubMed: 14570037]
18. Contino M, Capparelli E, Colabufo NA & Bush AI Editorial: The CB2 Cannabinoid System: A New Strategy in Neurodegenerative Disorder and Neuroinflammation. *Front Neurosci* 11, 196 (2017). [PubMed: 28428743]
19. Lunn CA et al. Biology and therapeutic potential of cannabinoid CB2 receptor inverse agonists. *Br J Pharmacol* 153, 226–39 (2008). [PubMed: 17906679]
20. Corey EJ General methods for the construction of complex molecules. *Pure and Applied Chemistry* 14, 19–38 (1967).
21. Baell JB & Holloway GA New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53, 2719–40 (2010). [PubMed: 20131845]
22. Li X et al. Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* 176, 459–467 e13 (2019). [PubMed: 30639103]
23. Kroeze WK et al. PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat Struct Mol Biol* 22, 362–9 (2015). [PubMed: 25895059]
24. Gaulton A et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40, D1100–7 (2012). [PubMed: 21948594]
25. Xing C et al. Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex. *Cell* 180, 645–654 e13 (2020). [PubMed: 32004460]
26. Wei L, Surma M, Shi S, Lambert-Cheatham N & Shi J Novel Insights into the Roles of Rho Kinase in Cancer. *Archivum Immunologiae et Therapiae Experimentalis* 64, 259–278 (2016). [PubMed: 26725045]
27. Chin VT et al. Rho-associated kinase signalling and the cancer microenvironment: novel biological implications and therapeutic opportunities. *Expert Rev Mol Med* 17, e17 (2015). [PubMed: 26507949]

28. Baker M Fragment-based lead discovery grows up. *Nat Rev Drug Discov* 12, 5–7 (2013). [PubMed: 23274457]
29. Schulz MN & Hubbard RE Recent progress in fragment-based lead discovery. *Curr Opin Pharmacol* 9, 615–21 (2009). [PubMed: 19477685]
30. Davis BJ & Hubbard RE Fragment-Based Ligand Discovery. in *Structural Biology in Drug Discovery* 79–98 (2020).
31. Zheng Z et al. Structure-Based Discovery of New Antagonist and Biased Agonist Chemotypes for the Kappa Opioid Receptor. *J Med Chem* 60, 3070–3081 (2017). [PubMed: 28339199]
32. de Graaf C et al. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J Med Chem* 54, 8195–206 (2011). [PubMed: 22007643]
33. Katritch V et al. Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J Med Chem* 53, 1799–809 (2010). [PubMed: 20095623]
34. Chen Y & Shoichet BK Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol* 5, 358–64 (2009). [PubMed: 19305397]

References for Methods only

35. Abagyan RA, Orry A, Raush E, Budagyan L & Totrov M ICM User's Guide and Reference Manual. 3.9 edn (MolSoft LLC, La Jolla, CA, 2021).
36. Bogolubsky AV et al. A One-Pot Parallel Reductive Amination of Aldehydes with Heteroaromatic Amines. *ACS Combinatorial Science* 16, 375–380 (2014). [PubMed: 24956115]
37. Savych O et al. One-Pot Parallel Synthesis of 5-(Dialkylamino)tetrazoles. *ACS Combinatorial Science* 21, 635–642 (2019). [PubMed: 31437394]
38. Katritch V, Rueda M & Abagyan R Ligand-guided receptor optimization. *Methods Mol Biol* 857, 189–205 (2012). [PubMed: 22323222]
39. Gatica EA & Cavasotto CN Ligand and decoy sets for docking to G protein-coupled receptors. *J Chem Inf Model* 52, 1–6 (2012). [PubMed: 22168315]
40. Bottegoni G, Kufareva I, Totrov M & Abagyan R Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J Med Chem* 52, 397–406 (2009). [PubMed: 19090659]
41. Enamine. <https://enamine.net/library-synthesis/real-compounds/real-compound-libraries>. (2020).
42. Nikas SP et al. Probing the carboxyester side chain in controlled deactivation (–)- δ (8)-tetrahydrocannabinols. *J Med Chem* 58, 665–81 (2015). [PubMed: 25470070]
43. Nikas SP et al. Novel 1',1'-chain substituted hexahydrocannabinols: 9 β -hydroxy-3-(1-hexyl-cyclobut-1-yl)-hexahydrocannabinol (AM2389) a highly potent cannabinoid receptor 1 (CB1) agonist. *J Med Chem* 53, 6996–7010 (2010). [PubMed: 20925434]
44. Jacobs M et al. The Structure of Dimeric ROCK I Reveals the Mechanism for Ligand Selectivity *. *Journal of Biological Chemistry* 281, 260–268 (2006). [PubMed: 16249185]
45. Anastassiadis T, Deacon SW, Devarajan K, Ma H & Peterson JR Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* 29, 1039–45 (2011). [PubMed: 22037377]

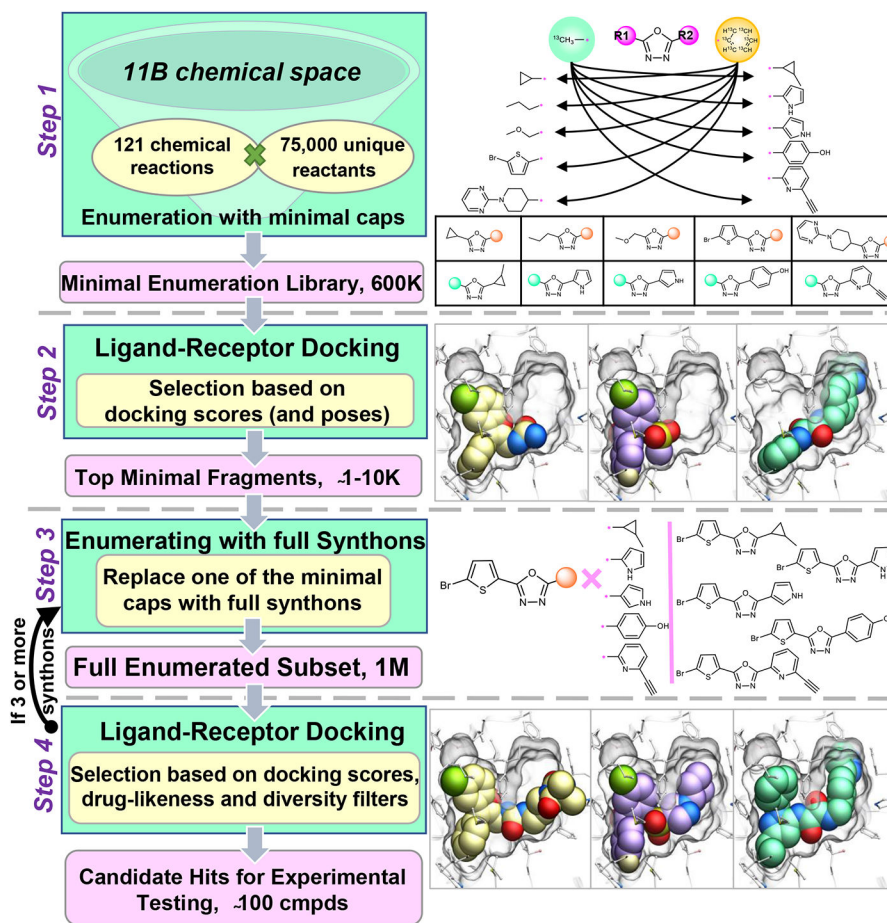


Figure 1. V-SYNTHES approach modular screening of Enamine REAL Space. The flowchart on the left shows a general overview of the 4-step algorithm, the panels on the right illustrate examples for each step.

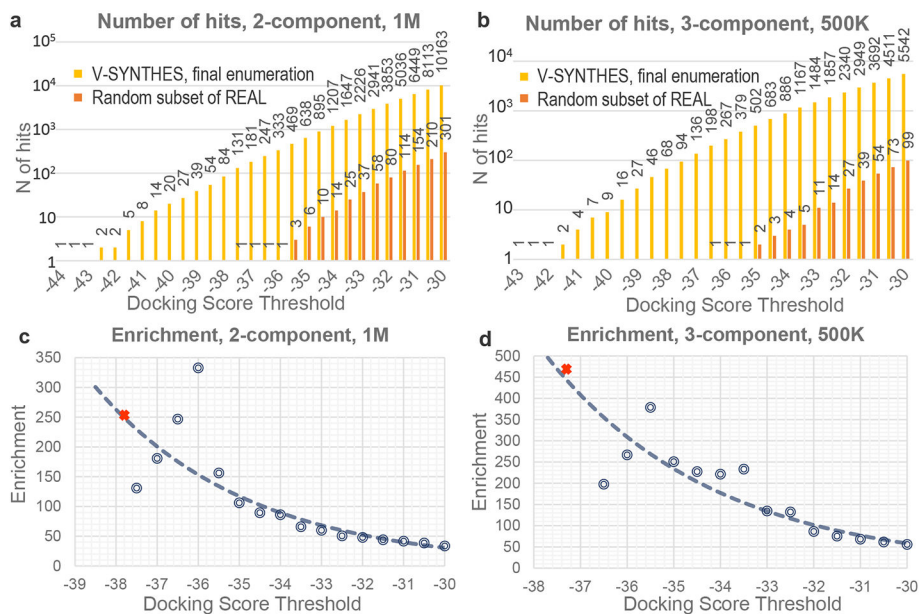


Figure 2. Assessment of VLS computational performance for V-SYNTHES and standard VLS. (a,b) The number of hits at each score threshold from V-SYNTHES and standard VLS (c,d) Enrichment in V-SYNTHES vs. Standard VLS at different score thresholds, where the red X-markings represent thresholds that yield 100 hits in 2 component (c) and 3-component (d) cases.

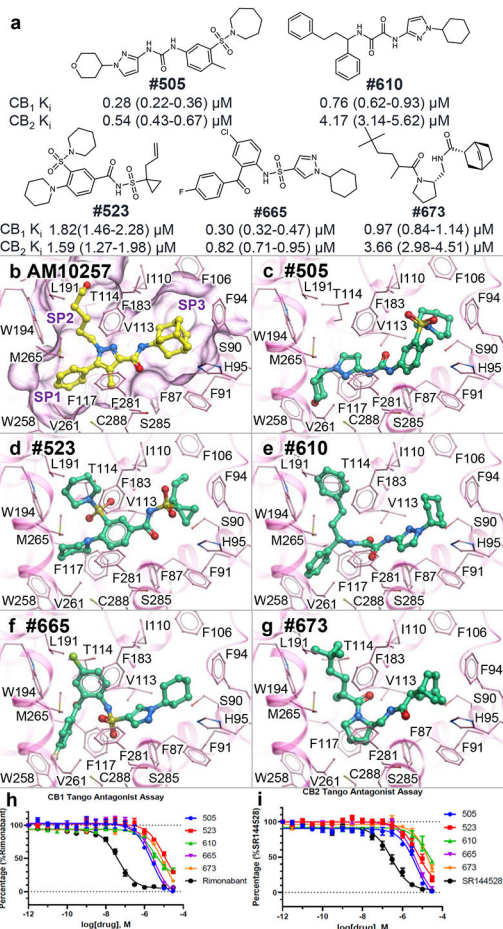


Figure 3. Top five CB₂ hits identified by V-SYNTHES.

(a) Chemical structures and measured antagonist potencies for CB₁ and CB₂ receptors. (b) Crystal structure of CB₂ receptor with AM10257 and predicted binding poses for hit compounds (c) 505, (d) 523, (e) 610, (f) 665, and (g) 673 in CB₂ receptor, respectively. Key subpockets of the binding pocket marked as SP1, SP2, and SP3. (h-i) Concentration-response curves for the top antagonists in β-arrestin recruitment Tango assays at CB₁ (h) and CB₂ (i) receptors. The assays were carried out in the presence of 100 nM (EC₈₀) of the dual CB₁/CB₂ agonist CP55,940. The compounds rimonabant (h) and SR144528 (i) served as positive controls. The data are presented as mean ± SEM from n=3 independent experiments, each run carried out in triplicate.

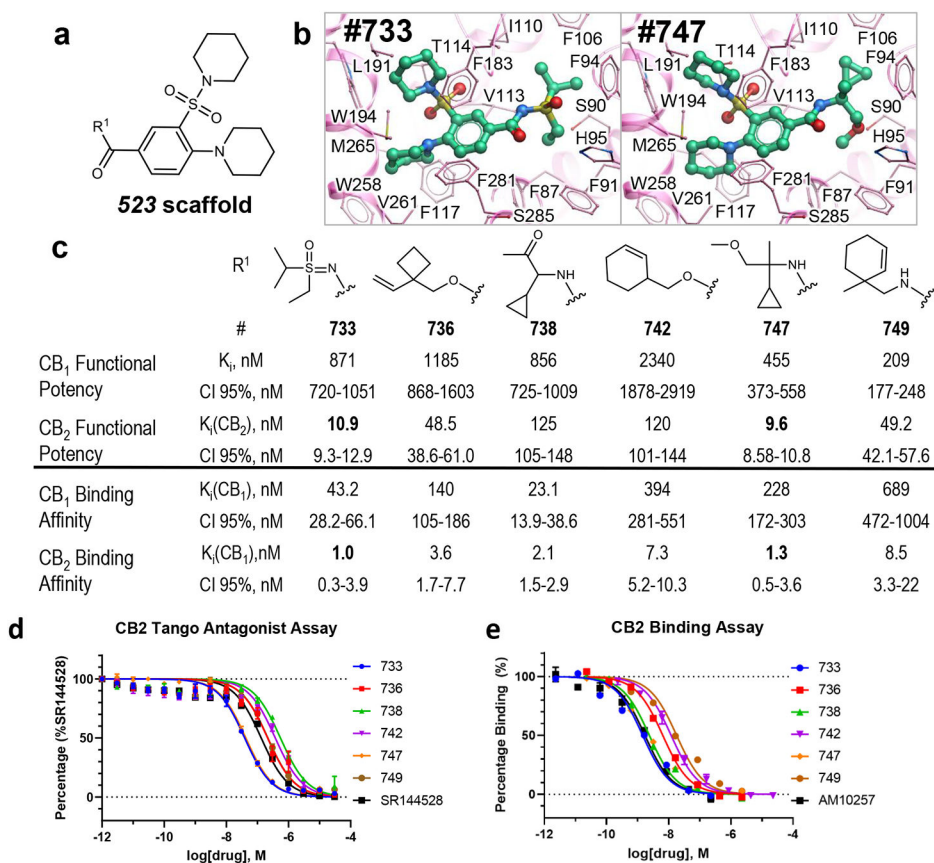


Figure 4. Selection and characterization of the best analog series for CB₂ hits from V-SYNTHES screening.

(a) Chemical scaffold for the **523** antagonist analogs (b) Predicted binding poses of the best two analogs **733** and **747** in CB₂ pocket. (c) Measured antagonist potencies and binding affinities for the best six analogs of **523**. (d) Dose-response curves for the best six analogs tested in functional β -arrestin recruitment Tango assays at CB₂, with SR144528 as a positive control. (e) Dose-response curves for the best six analogs at CB₂ tested in radioligand binding assay, with compound AM10257 as a positive control. Both the Tango and binding assay data points are presented as mean \pm SEM with n=3 independent experiments, each repeat carried out in triplicate.