Subpocket similarity-based hit identification for challenging targets: Application to the WDR domain of LRRK2

Merveille Eguida, Guillaume Bret, François Sindt, Fengling Li, Irene Chau, Suzanne Ackloo, Cheryl Arrowsmith, Albina Bolotokova, Pegah Ghiabi, Elisa Gibson, et al.

Subpocket similarity-based hit identification for challenging targets: Application to the WDR domain of LRRK2.

Merveille Eguida,[†,#,*] Guillaume Bret,[†] François Sindt,[†] Fengling Li,[‡] Irene Chau,[‡] Suzanne Ackloo,[‡] Cheryl Arrowsmith,[‡,&] Albina Bolotokova,[‡] Pegah Ghiabi,[‡] Elisa Gibson,[‡] Levon Halabelian,[‡,&] Scott Houliston,[&] Rachel J. Harding,[‡,%] Ashley Hutchinson,[‡] Peter Loppnau,[‡] Sumera Perveen,[‡] Almagul Seitova,[‡] Hong Zeng,[‡] Matthieu Schapira[‡,%] and Didier Rognan[†,*]

[†] Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch

[‡] Structural Genomics Consortium, University of Toronto, Toronto, ON, M5G 1L7, Canada

[&] Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

[%] Department of Pharmacology & Toxicology, University of Toronto

[#] Current address: Amgen Research Copenhagen, 2100 Copenhagen, Denmark

* To whom correspondence should be addressed (phone: +33 3 68 85 42 35, fax: +33 3 68 85 43 10, email: meguida@amgen.com, rognan@unistra.fr,)

ABSTRACT

We herewith applied a generic hit identification method (POEM) for *a priori* difficult targets of known three-dimensional structure, relying on the simple knowledge of physicochemical and topological properties of a user-selected cavity. Searching for local similarity to a set of fragment-bound protein microenvironments of known structure, a point cloud registration algorithm is first applied to align known subpockets to the target cavity. The resulting alignment permits then to directly pose the corresponding seed fragments in a target cavity space not typically amenable to classical docking approaches. Last, linking potentially connectable atoms by a deep generative linker enables full ligand enumeration. When applied to the WD40 repeat (WDR) central cavity of leucine-rich repeat kinase 2 (LRRK2), an unprecedented binding site, POEM was able to quickly propose 94 potential hits, five of which were subsequently confirmed to bind *in vitro* to LRRK2-WDR.

INTRODUCTION

In absence of any ligand information but the three-dimensional structure of a target of interest, docking-based virtual screening of compound libraries is the computational method of choice to identify the very first ligands (hits)[1] before hit-to-lead optimization. Docking is usually restricted to a cavity/pocket of interest, and many algorithms are now available to automatically detect cavities at the surface of macromolecules using either geometric, energetic and more recently data-driven approaches.[2-3] Once detected, various machine learning and deep learning models[4] can be trained on topological and physicochemical pocket descriptors (e.g. size, curvature, hydrophobic/hydrophilic balance) to predict their structural druggability, in other words the probability to host a high-affinity low molecular-weight ligand. This approach nevertheless requires that the target exhibits a druggable cavity at its surface, in other words a pocket with physicochemical (good hydrophilic vs. hydrophobic balance) and topological (e.g. good accessibility and buriedness) properties suitable to accommodate a drug-like compound.[5, 2] For pockets not having such characteristics (e.g. ion channel pores[6] or flat protein-protein/peptide interfaces[7]), docking methods are expected to be less appropriate, notably because of the multiplicity of potential docking solutions, and the difficulty to properly rank them and prioritize potential hits. As an alternative, one may take advantage of the pocket similarity principle stating that *similar pockets binds similar ligands* and simply test known ligands of closely related pockets for binding to the query cavity.[8-9] To be efficient, the pair of pockets to compare need to be globally similar in their total shape and physicochemical properties. In case of partial similarity (just a part of the query cavity is similar to a part of a remote pocket), transferring ligands from one pocket to another one is much more difficult and at best usually restricted to shared fragments.[10-11] To address this difficult issue, we recently proposed a novel computational approach (POEM: Pocket Oriented Elaboration of Molecules)[12] that generates ligands from the simple knowledge of a target cavity, by automatically linking fragments originally bound to remote proteins exhibiting only local similarities to the query cavity. When applied to a protein kinase (CDK8), POEM was able to quickly generate a single-

digit nanomolar inhibitor in two Design-Make-Test-Analyze (DMTA) cycles[13] and 43 synthesized ligands.[12]

Although the POEM idea falls within the concept of target-based de novo drug,[14-15] it differs from existing methods by a combination of several aspects: (i) no reference ligand is required for the target while some methods (e.g., BREED,[16] KinFragLib[17]) rely on reference protein-ligand complexes for molecular hybridization, (ii) pairs of fragments are directly used for elaboration, as opposed to (grid-based) sampling of atoms as in BUILDER,[18] CONCEPTS[19] or Ramensky et al.,[20] (iii) the fragment templates are derived from existing protein-ligand complexes in their X-ray conformation, instead of using a library of template fragments as in LUDI,[21] LigBuilder,[22] or FastGrow,[23] (iv) fragments are positioned according to the similarity of their subpocket to the target cavity and are not scored by any energy criteria (e.g., GroupBuild,[24] LUDI)[21], (v) the fragments linking is based on a 3D-constrained variational autoencoder to generate potential linker graphs, instead of strict topological generators guided by explicit bond and torsion angle ranges.[25] The closest implementations to POEM are the work by Moriaud et al.[26] and Durrant et al.,[27] suggesting building block fragments to link on the basis of their environment similarity with the target site, albeit with a different site representation and comparison algorithms. Moreover, the latter methods do not enumerate fully connected molecules from the position of seed fragments.

Whether the POEM workflow could be applied to a challenging target is herein presented, as part of the CACHE (Critical Assessment of Computational Hit-finding Experiments) challenge #1.[28] The CACHE challenge aims to publicly benchmark the ability of computational methods to predict hits for relevant targets by confirming predictions with experimental validation. For this first edition, the WD40 repeat (WDR) domain of the human leucine-rich repeats kinase 2 (LRRK2)[29] was chosen by the organizers. The WDR domain, a β-propeller of seven blades, has been shown to mediate LRRK2 protein-protein interactions with microtubules and vesicle trafficking in neurons.[30] Sporadic or inherited mutations in the *LRRK2* gene are commonly associated with Parkinson's disease.[31] To date, no LRRK2-targeting drug

has reached the market, current candidates being either small molecules inhibiting the kinase domain (mostly ATP-competitive inhibitors) or nucleic acids (antisense oligonucleotides siRNAs, shRNAs, microRNAs and aptamers) binding to the mRNA of the *LRRK2* gene and lowering the cellular level of LRRK2 .[32] The goal of the challenge is to propose low molecular-weight compounds binding to the WDR domain and therefore prevent the formation of LRRK2 filaments in cells. Due to the absence of known LRRK2-WDR ligands, and the nature of the WDR inner pore, this challenge provides a unique opportunity to test the POEM method in a ligand-unbiased prospective manner.

MATERIAL AND METHODS

sc-PDB database of druggable protein-ligand complexes (version v.2022). A total of 184,202 PDB entries (mmcif format) were locally mirrored on December 10, 2021. Each file was browsed to extract annotations and atomic coordinates and converted to pdb format with an in-house python script. For nuclear magnetic resonance (NMR) structures, only atoms from the representative first model were retained. In case of alternate atom locations, only the first location was kept. Hydrogen atoms were added with Protoss v.4.0[33] to optimize both protonation and tautomeric states, to add missing hydrogens, to select the best orientation of ambiguous amino acid side chains (Asn, Gln, His) and optimize the location of water hydrogen atoms. IChem v5.2.9[34] was next used to detect the number of hydrogen bonds between each water molecule and surrounding non-water atoms. Water molecules exhibiting less than 3 interactions with non-water molecules were removed. Ligands were defined from the annotations previously extracted from the mmcif file and classified as follows: small nucleotides (< 4-mer), short peptides (< 9-mer), and organic compounds (any remaining molecule). Organic ligands were further filtered with tabu lists excluding ions, metals, prosthetic groups and water. Organics compounds of unknown type were considered as potential ligands.

The binding site was defined as any residue (protein, co-factor, ion, water) with at least one heavy atom within 6.5 Å of any ligand heavy atom. Altogether, 148,784 cavities were detected from the bound ligand using the VolSite[35] module of IChem and further filtered according to their structural druggability (druggability probability ≥ 0),[35] number of non-covalent interactions with the bound ligand (≥ 5 with at least one being polar: hydrogen-bond, ionic bond or π- π aromatic), density (ratio of VolSite cavity points to binding site heavy atoms ≥ 0.3), and buriedness of the bound ligand (buried surface area ≥ 0.5 x total surface). Finally, all remaining cavities of each PDB entry were sorted according to the molecular weight of the bound ligand. Duplicates (same ligand bound to two dissimilar cavities of the same PDB entry) were removed by computing pairwise cavity similarity searches with Shaper v.2[35] and using a similarity score (color Tversky) of 0.5 as similarity threshold. In case two cavities of the same PDB entry were found similar (similarity ≥ 0.5), the one with the largest ligand was retained. The

final sc-PDB release (v.2022) contains 37,922 unique druggable cavities representing 30,109 PDB entries.

sc-PDB fragments and subpockets. Fragments and subpockets were prepared as previously described,[36, 12] from protein-ligand complexes of the v.2022 sc-PDB release. Ligands were fragmented in the 3D space of their bound target using two fragmentation schemes: (i) the default IChem fragmentation protocol,[36] and (ii) a novel 3D implementation of RECAP fragmentation.[37] Caution was given to remove duplicates (same fragments bound to the same sc-PDB entry) resulting from the two fragmentation schemes. Exit dummy atoms resulting from the fragmentation were converted into hydrogen atoms with SYBYL-X v.2.1.1.[38] Each subpocket was finally defined as a VolSite point cloud of 1.5 Å grid space and within 4 Å of the fragment center of mass, discarding those with less than 3 cavity points. The sc-PDB subpocket-fragment database consists of 107,828 unique entries.

Protein structure preparation. The dimeric structure of the LRRK2 WDR domain[39] displaying the highest resolution out of all available entries (PDB ID 6DLO) was downloaded from the RCSB PDB website.[40] The structure was protonated with Protoss v.4.0[33] and converted into MOL2 format with SYBYL-X v.2.1.1 while discarding the chain B of the dimer. No water molecule was retained. The inner cavity of the WDR domain was represented as a point cloud using standard grid settings (1.5 Å resolution, 20.0 Å-long edges) of the VolSite[35] module implemented in IChem v.5.2.9.[34]

Subpockets to query cavity alignment. sc-PDB subpockets were compared to the LRRK2-WDR cavity with ProCare v.0.1.2,[10] using the color c-FH descriptor and corresponding scoring scheme, and finally retaining subpockets yielding a similarity score above 0.47, a threshold previously shown to optimally discriminate known similar from known dissimilar pocket pairs.[10] The alignment matrices obtained were next applied to the corresponding fragments to pose them in the target cavity.

Fragments pharmacophoric points. Fragments heavy atoms were converted into the same set of pharmacophoric points as VolSite cavities with in-house scripts following these rules: OD1 for negatively charged atoms; NZ for positively charged atoms; OG for hydroxyl groups; O for H-bond acceptors and N for H bond donors; CA for hydrophobic atoms (carbon, sulfur, chlorine, bromine, and iodine not connected to a hetero atom). Aromatic groups were represented by features (CZ) placed at the center of the ring. Multiple features were allowed for a single position, when applicable (e.g., NZ and N to encode ammonium moiety).

Linking of ProCare fragments into POEM hits. Linking of selected fragments occurs in the following steps:

- (i) definition of connectable fragments and atoms. Two subpockets 'centers' (*top* and *bottom*) were automatically defined as the set of cavity points in LRRK2-WDR successfully aligned to more than 25% of sc-PDB subpocket hits. One is located around Y2249 (N-terminal bottom side), the second area is at the opposite side close to M2301 (C-terminal top side). Selected fragments were clustered into the nearest center considering the median of pairwise fragments atoms-cluster points minimum Euclidian distances. A '*middle'* cluster category was assigned to the fragments laying in-between the clusters (difference of median distance < 7.5 Å). Upon visual inspection of the centers, connectable subpockets and subsequently their respective fragments were associated in a combinatorial manner (*top-middle, top-top, bottom-middle, bottom-bottom*). For each pair of connectable fragments, heavy-hydrogen atoms were used as exit vectors and pairs of linkable atoms were finally registered when they face each other with a distance and dihedral degree of freedom constrained by a projected cone as previously described (**Supporting Figure S1**).[12] Overlapping fragments, defined as having three or more pairwise atom distances lower than 1.3 Å, were discarded.
- (ii) linking. Pairs of connectable atoms were linked using default parameters of the deep generative DeLinker software.[41]

- (Iii) post-processing of generated molecules. Duplicated molecules for the same pair of connectable atoms were removed. In case of a failure to successfully connect both ends of the linker to the two seed fragments, the corresponding incomplete ligand was discarded.
- (iv) filtering. Molecules were indexed and filtered for drug-likeness using OpenEye Filter v.3.0.1.2[42] and in-house custom rules **(Supporting Table S1**).

Commercial availability of POEM hits. The Enamine REAL diverse set of 38 million commercially available molecules was downloaded[43] and filtered for drug-likeness (**Supporting Table S1**) with OpenEye Filter v.3.0.1.2 yielding 24 million drug-like molecules. They were compared to POEM virtual hits using Morgan2 fingerprints computed with RDKit v.2019.03.4.0.[44] REAL molecules exhibiting a Tanimoto similarity higher than 0.70 were retained and clustered according to their ECFP4 fingerprints by the R-Agnes method implemented in PipelinePilot v.22.1.0.2935.[45]

Redocking of experimentally confirmed hits. 2D sketches of experimentally confirmed hits were obtained with BioviaDraw v.21.1,[45] converted in 3D MOL2 files with Corina v3.40,[46] and docked into the LRRK2-WDR cavity using standard parameters of the FitDock algorithm.[47] For each hit, the two originally selected ProCare fragment poses were just used as templates for the multi-features alignment to constrain the docking.

Hit expansion. Maximum common substructure (MCS) similarity of validated hits (converted from mol2 to SMILES strings using Open Babel v.3.1.0)[48] to 36 billion REAL space ligands (version REALSpace_36bn_2023-03.space)[49] was computed with SpaceMACS v.0.9.2,[50] to save the top 20 REAL space compounds ranked by decreasing MCS-Tanimoto similarity value.

Protein expression and purification. DNA fragments encoding LRRK2 residues (T2124- E2527) and (T2141- E2527) were cloned into pFastBac HTA donor plasmid downstream of a His-tag or into pFBD-BirA expression vector, a derivative of Invitrogen pFastBac Dual vector for in-cell biotinylation (https://www.thesgc-dev.org/sites/default/files/toronto_vectors/pFB-BirA.pdf), respectively. The resulting plasmid was transformed into DH10Bac™ Competent E. coli (Invitrogen) to obtain recombinant viral bacmid DNA, followed by a baculovirus generation for protein production in Sf9 insect cells. For in-cell biotinylation, D-biotin was added at the final concentration of 10 µg/mL during protein expression. The cells were harvested by centrifugation (2500 rpm for 10 mins at 10°C), 72-96 hours post-infection with well-developed signs of infections and 70-80 % viability as previously described.[51] Harvested cells were resuspended in 20mM Tris-HCl, pH 7.5, 500mM NaCl, 5mM imidazole and 5% glycerol, 1X protease inhibitor cocktail (100 X protease inhibitor stock in 70% ethanol (0.25mg/ml Aprotinin, 0.25mg/ml Leupeptin, 0.25mg/ml Pepstatin A and 0.25mg/ml E-64) or Pierce™ Protease Inhibitor Mini Tablets, EDTA-free. The cells were lysed chemically by addition of 1mM PMSF, 1mM TCEP, 0.5% NP40 and benzonase (in-house) followed by sonication at frequency of 7.0 (5" on/7" off) for 5 min (Sonicator 3000, Misoni). The crude extract was clarified by high-speed centrifugation (60 min at 14000 rpm at 10˚C) by Beckman Coulter centrifuge. The clarified lysate was loaded onto open columns containing pre-equilibrated Ni-NTA resin (Sigma Aldrich). The column was washed and eluted by running 20mM Tris-HCl, pH 7.5, 500mM NaCl, 5% glycerol, containing 5mM, 15mM and 250mM imidazole, respectively. The eluted proteins were then supplemented with 2mM TCEP. The His- and Avi-tagged protein was then further purified by size-exclusion chromatography on a Superdex200 16/600 using an ÄKTA Pure (Cytiva) after the column was equilibrated with 50mM Tris-HCl pH 7.5, 300mM NaCl, 2mM TCEP.

For the His-tagged protein, the tag was cleaved after elution using tobacco etch virus protease (TEV) overnight while the protein was dialyzed against 20mM Tris-HCl, pH 7.4, containing 300mM NaCl, 2mM TCEP. The protein was then loaded on equilibrated Ni-NTA resin for reverse affinity to remove His-tagged TEV enzyme and the uncut His-tagged proteins. The purity and size of the cut protein was

confirmed on SDS-PAGE gel and mass spectrometry, respectively and the pure protein was concentrated and flash frozen.

Surface Plasmon Resonance (SPR). The binding affinity of compounds was assessed by surface plasmon resonance (SPR, Biacore™ 8K, Cytiva Inc.) at 25 °C. Biotinylated LRRK2 (2141-2527aa) was captured onto flow cells of a streptavidin-conjugated SA chip at approximately 5,000 response units (RU) (according to manufacturer's protocol). Compounds were dissolved in 100% DMSO (30 mM Stock) and diluted to 10 mM before serial dilutions were prepared in 100% DMSO (dilution factor of 0.33 was used to yield 5 concentrations). For SPR analysis, serially titrated compounds were diluted 1:50 in HBS–buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.01% Tween-20) to a final concentration of 2% DMSO. Experiments were performed using the same buffer containing 2% DMSO and multi-cycle kinetics with a 60 s contact time and a dissociation time of 120 s at a flow rate of 40 µL/min. Steady state curve fittings and K_D value calculations were done with a 1:1 binding model using the Biacore Insight Evaluation Software (Cytiva Inc).

Differential Scanning Fluorimetry (DSF). LRRK2 was diluted to 0.1 mg/ml in buffer (100 mM Hepes pH 7.5, 150 mM NaCl) in the presence of 5x SYPRO Orange dye (Life Technologies, S-6650) and serially titrated compounds from DMSO stocks, followed by a 50x dilution into buffer, resulting in 2% final DMSO in a white polypropylene 384-well plate (Axygen, PCR-384-LC480-W). DSF was performed in a Light Cycler 480 II (Roche Applied Science, Penzberg, Germany) using a 4°C/min temperature gradient from 20°C to 95°C. Data points were collected at 0.5°C intervals. DSF data was fitted to a Boltzmann sigmoid function and T_m values were determined as previously described.[52]

Dynamic Light Scattering (DLS). Compounds were serially diluted from DMSO stocks, followed by a 50x dilution into filtered buffer (10 mM HEPES pH 7.4, 150 mM NaCl without detergent, 2% final

DMSO concentration). The resulting samples were then distributed into 384-well plates (black with a clear bottom, Corning 3540), with each well 20µl. The sample plate was centrifuged at 3500 rpm for 5 minutes before loading into DynaPro DLS Plate Reader III (Wyatt Technology). Normalized intensity (kCnt/s) and laser power (%) were used as indicators for estimating compound aggregation and solubility. Intensity above $1x10^{+6}$ Cnt/s was considered to be from the aggregated form, while decreased laser power indicates poor solubility.

^{19}F-NMR spectroscopy. The binding of fluorinated compounds was assayed by looking for the broadening and/or perturbation of ^{19}F resonances upon addition of LRRK2 (at protein to compound ratios of 0.5:1 to 4:1) in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, 1.8 mM KH_2PO_4 pH7.4, and with 5% D_2O). 1D-^{19}F spectra were collected at 298K on a Bruker Avance™ III spectrometer, operating at 600 MHz, and equipped with a QCI probe. Two to four thousand transients were collected with an acquisition period of 0.2 s, over a sweep width of 150 ppm, a relaxation delay of 1.5 s, and using 90° pulses centered at -120 ppm. The concentration of the compounds in both reference and protein-compound mixtures was 5-10 µM. TFA (20 µM) was added as an internal standard for referencing. Prior to Fourier transformation, an exponential window function was applied (lb = 1 to 3) to the FID. All processing was performed at the workstation using the software Topspin3.5 (Bruker).

RESULTS AND DISCUSSION

LRRK2-WDR structure and cavity definition. The target WDR domain is located at the extreme C-terminus (residues 2142-2497) of LRRK2 (**Figure 1A**) and described, at the launch of the challenge, by several experimentally-determined structures (**Figure 1B**) in the Protein Data Bank (PDB): a 2.6 Å-resolution X-ray structure of the isolated WDR domain (PDB 6DLO) in a dimeric state,[39] a 3.5 Å cryo-EM structure of the LRRK2 C-terminal domain (PDB 6VNO),[30] a 3.8 Å cryo-EM structure of the full length LRRK2 G2019S monomer (PDB 7LI3),[29] and a 3.7 Å cryo-EM structure of the full length LRRK2 dimer (PDB 7LHT).[29] The 3D structure of the WDR domain is structurally conserved when comparing cryo-EM structures to the high-resolution X-ray structure (RMSD of 1.2 Å on heavy atoms). We therefore chose the latter structure for POEM application, acknowledging the 49 missing residues in the 6 outer loops (we hence applied the same workflow to the 7LHT structure with resolved loops for checking obtained results). The inner cavity, detected and characterized with the in-house VolSite[35] algorithm (**Figure 1C**) exhibits a cylinder shape of ca. 22 Å long and 8 Å wide at its center, fully accessible from both sides. It is described by 361 pharmacophoric points, represented in majority by hydrophobic and hydrogen-bond donor properties (**Figure 1D**). Thirty-three residues delimit the cavity, with polar amino acids located at both rims of the pore (E2153 and R2456 at one side, Y2249, S2304, T2460 at the other side) and almost only hydrophobic residues in the center. The buriedness of the cavity, estimated by VolSite is relatively low (69%), notably because of the two wide apertures at both ends of the cylindric shape. Despite the cavity is predicted to be structurally druggable (druggability=1.40) by our machine learning (ML) model,[35] the prediction has to be taken with great caution because of the peculiar shape of the pore, very different from protein cavities on which the druggability ML model has been trained.

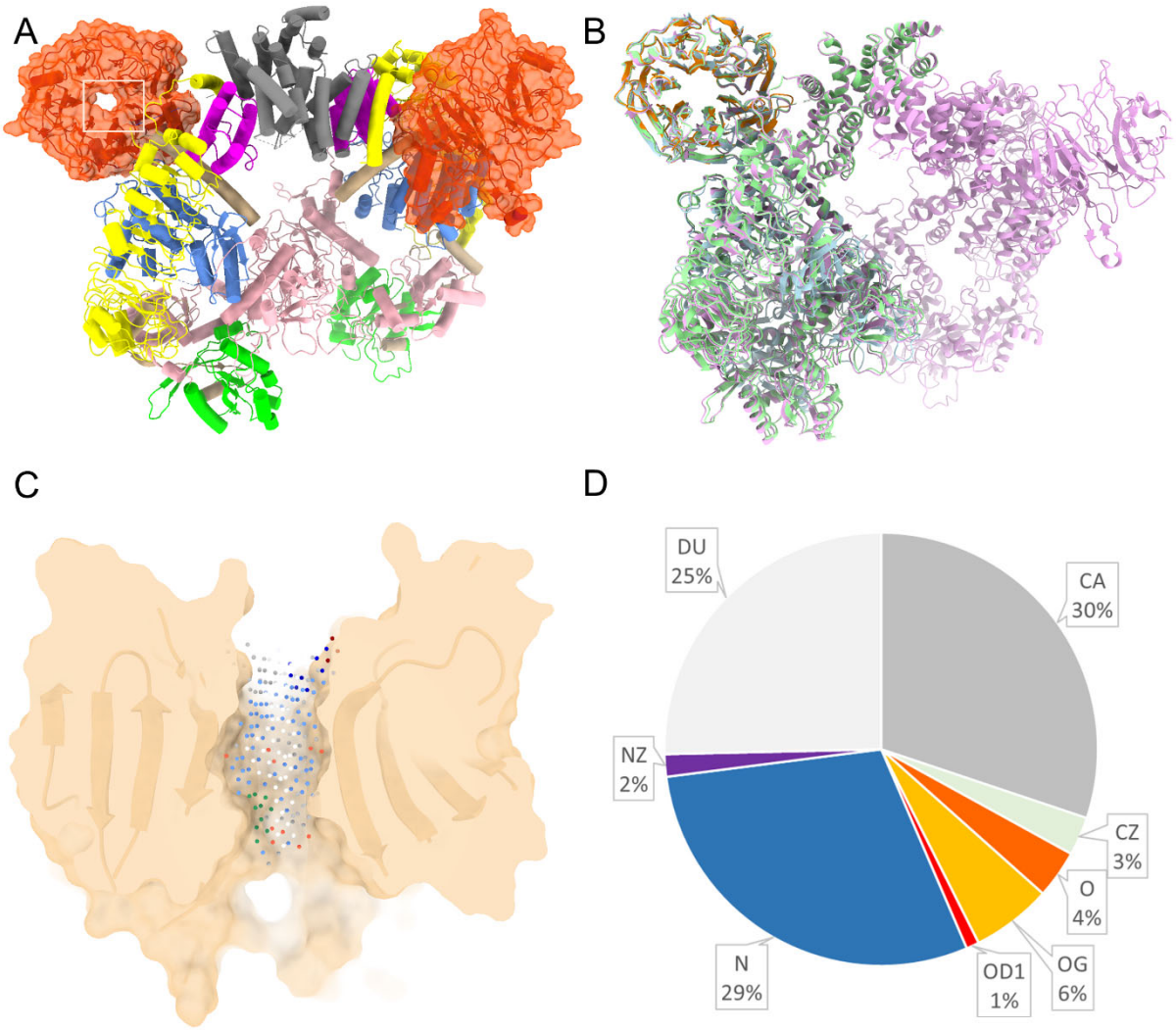

Figure 1. Preparation of the LRRK2-WDR target for the POEM workflow. A) Cryo-electron microscopy structure (PDB 7LHT) of human full length LRRK2 dimer (ARM domain: grey, ANK domain: magenta, LRR domain: yellow, ROC domain: green, COR domain: light pink; KIN domain: cornflower blue, WDR domain: orange). The WDR domain is shown as a solid transparent surface with its inner pore delimited by a rectangle. **B)** Structural alignment of four LRRK2-WDR structures: X-ray structure of the isolated WDR domain (PDB 6DLO, orange), LRRK2 C-terminal monomer (PDB 6VNO, cyan), LRRK2 full-length monomer (PDB 7LI3, green), LRRK2 full length dimer (PDB 7LHT, magenta). **C)** VolSite point cloud representation of the inner cavity of the WDR monomer (PDB ID 6DLO, chain A). Each cavity point is colored according to pharmacophoric properties (hydrophobic, grey; aromatic, green; hydrogen-bond acceptor, donor and acceptor or negatively charged, red; hydrogen-bond donor or positively charged, blue; no property, white) of an ideal ligand filling the cavity. The WDR domain (tan solid surface) has been clipped along the z-axis to better reveal the inner cavity. **D)** Distribution of pharmacophoric properties among cavity points (CA, hydrophobic; CZ, aromatic; O, hydrogen-bond acceptor; OG: hydrogen-bond acceptor and donor; OD1, negative ionizable; N: hydrogen-bond donor; NZ, positive ionizable; DU, property assigned to every cavity point for which the closest protein atom is farther than 4.0 Å).

Seed fragments selection. 107,828 subpockets, extracted from the sc-PDB database of druggable protein-ligand complexes,[53] were aligned to the LRRK2-WDR cavity using a point cloud registration algorithm (ProCare),[10] specifically designed to find the best possible alignments between two clouds of cavity points and estimate their local similarity.[11-12] Given the high proportion of polar features in the query cavity, alignment by the color descriptor (c-FH) was chosen to prioritize subpockets in three steps (**Figure 2**):

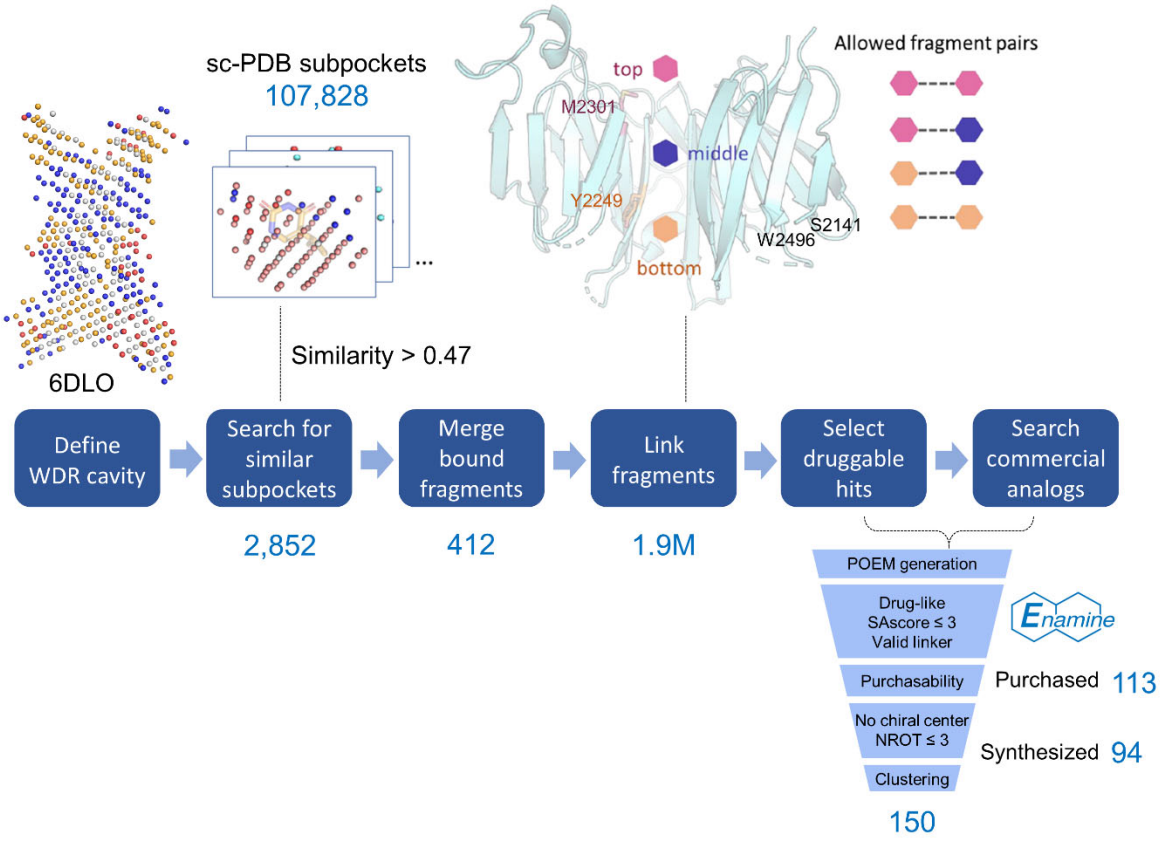

Figure 2. POEM workflow to generate hit candidates. After selection of the most similar sc-PDB subpockets, bound fragments are merged in the target cavity, prioritized to occupy three areas (top, middle, bottom) and linked by a deep generative algorithm to enumerate virtual hits that are pruned to yield on-demand synthesizable compounds. The N-terminal (S2141) and C-terminal (W2496) of the LRRK2-WDR domain are indicated for a precise location of top, middle and bottom fragment-anchoring zones.

(i) selection of 2,852 subpockets with a ProCare similarity score above 0.47, a threshold previously shown to optimally discriminate known similar from known dissimilar pocket pairs;[10] and merging of the corresponding fragments in the target cavity coordinate frame using the same rotation/translation matrix as that selected for aligning their subpockets; (ii) prioritization of 389 fragments (derived from

non-cofactors) for which at least 50% of their pharmacophoric features match, within a upper distance of 3.0 Å, a LRRK2-WDR cavity point of compatible pharmacophoric property, (iii) sampling and adding fragments from sc-PDB bound co-factors with respect to the observed co-factor-bound targets enrichment, yielding a final set of 412 fragments. The fragments were located along three delimited zones, from the upper side to the lower side of the inner core (**Figure 2**).

Fragment linking and virtual hit enumeration. To avoid a combinatorial explosion of plausible solutions, we restricted the number of linkable fragments according to preferred locations in the target cavity. The idea behind this filter is that LRRK2-WDR cavity points repeatedly utilized by ProCare for aligning sc-PDB subpockets are of high statistical value. Target cavity points that were aligned by at least 25% of productive alignments (consensus between 6DLO and 7LHT), delimited two preferential areas, one located around Y2249 (N-terminal bottom side), and a second area at the opposite side close to M2301 (C-terminal top side, **Figure 2**). These two areas are too far away (ca. 16 Å) to host seed fragments that may be directly connected by a short linker. We therefore defined a third middle zone to enable the enumeration of molecules of acceptable size according to four connecting possibilities: *top-top*, *top-middle, bottom-middle, and bottom-bottom* **(Figure 2)**.

In addition to rules implemented in our previous study,[12] we restricted pairs of connectable fragments in order to reach a cumulative size of 13 to 25 heavy atoms. Almost colinear and overlapping fragment planes are neither desirable since they would require distorted linkers. Subsequently, fragment pairs displaying a least 3 pairwise distances between 0 and 2 Å were discarded. These implementations clearly restricted the list of fragments to be linked. The DeLinker deep generative program[41] was next applied to the remaining seed fragments to enumerate 1.9 million SMILES strings that underwent a series of filters to reach a list of 150 hits requested by the CACHE challenge organizers: drug-likeness (**Supporting Table S1**), synthetic accessibility SAscore ≤ 3,[54] full enumeration (linker connected to both seed fragments), commercial availability of close analogs (2D similarity to commercial REAL drug-like

compounds ≥ 0.70, no chiral center, ≤ 6 rotatable bonds). Remaining hits were finally clustered according to ECFP4 fingerprints to yield 150 representative hits (**Figure 2**) out of which 94 compounds (**Supporting Table S2)** could indeed be synthesized by Enamine (Enamine Ltd, Kyiv, Ukraine) and further evaluated for binding to LRRK2-WDR.

Primary screening by surface plasmon resonance (Round 1). POEM hits were evaluated for binding to LRRK2-WDR domain by surface plasmon resonance (SPR) at 50 and 100 µM compound concentration. We here only report primary screening data at the lowest of the two concentrations. Out of the 94 tested compounds, 17 compounds exhibiting a measured versus expected signal (Response Unit or RU) ratio between 0.5 (significant binding) and 2.0 (above 2 indicates non-specific binding) were selected for follow-up experiments (**Figure 3**). Four compounds (#40, #56, #71, #86) were immediately discarded by visual inspection of the SPR signal and detection of inconsistencies (e.g. slow dissociation, poor sensorgrams, measured SPR signal RUmax higher than 2-fold of expected RUmax). In ligand titration experiments, dose-dependent binding signal was observed for four compounds (#13, #37, #80, and #97), and steady state kinetics analysis permitted to estimate dissociation constants (K_D) ranging from 23 to 45 µM (**Figure 3B**). The four hit candidates did not bind by SPR the PWWP domain of the nuclear receptor-binding SET domain-protein 2 (NSD2), an unrelated target. The percentage of binding to LRRK2-WDR remained weak for two compounds (#37 and #97), acceptable for one hit (#13), and excellent for hit #80 (**Figure 3B**). Interestingly, candidate hits share two common pharmacophoric features, an indole ring (common to compounds #13, #37), and a phenol moiety (common to hits #13, #80 and #97). An additional potential hit (#45, K_D ~ 50 µM, **Supporting Figure S2**) also bearing an indole ring was selected in a late analysis despite an imperfect SPR dose-response curve (no real saturation of the signal at high concentrations) for follow-up studies. At this stage, we must acknowledge that compounds that were experimentally tested are not identical to those prioritized by the POEM workflow (**Supporting Figure S2 and Table S2**), but closely related analogs available in REAL space, and usually differing by the nature of the linker. This is a clear known drawback of generative algorithms

that design compounds which often do not strictly intersect commercially available or synthesizable molecules. However, seed fragments selected by the initial subpocket similarity search are almost unchanged, with the exception of small substituent changes (#80 and #97) or bioisosteric replacements (#37; **Supporting Figure S2**). To get insights about putative binding modes of the confirmed hits and possible shared non-covalent interactions with LRRK2-WDR, docking to the X-ray structure of the WDR domain was undertaken, keeping the ProCare-predicted position of seed fragments as constraints.

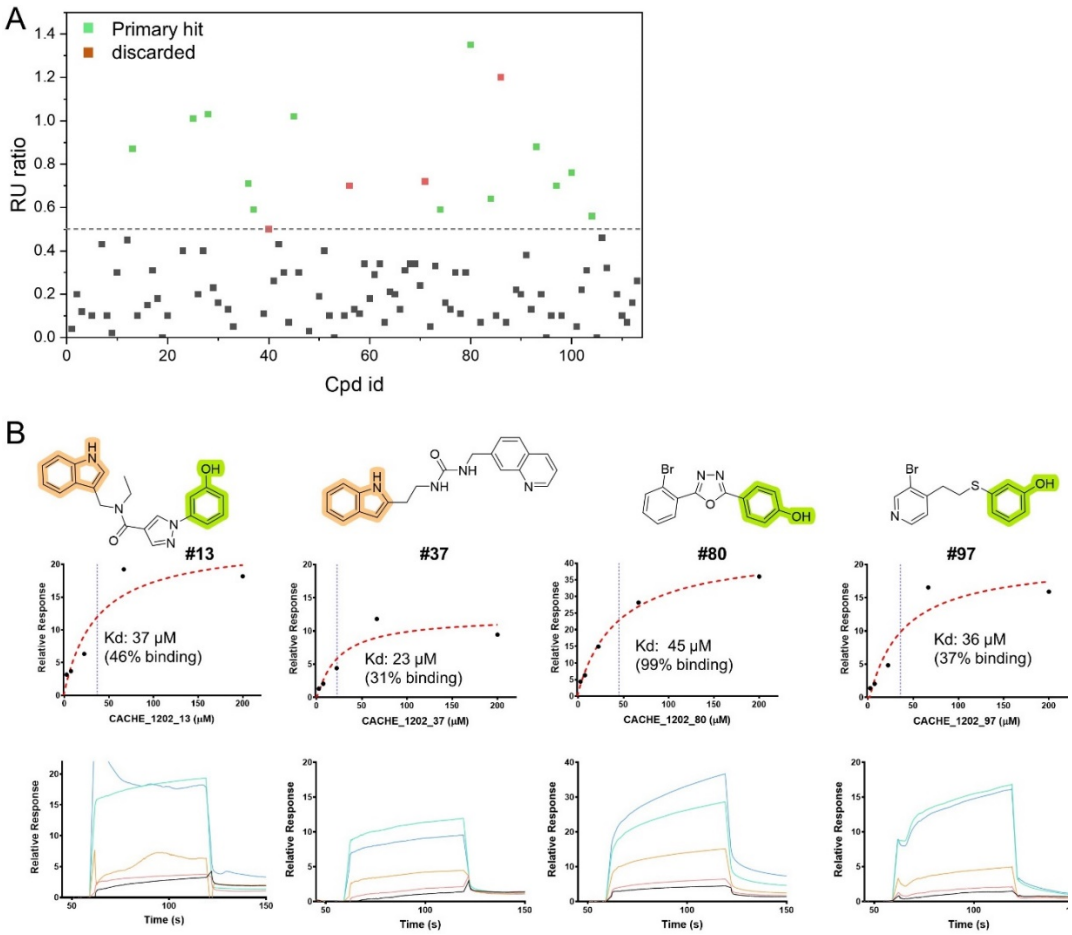

Figure 3. Primary screening by surface plasmon resonance. **A)** Binding of 94 POEM hits to the biotinylated LRRK2-WDR domain, at the single concentration of 50 µM. Primary hits (green squares) are selected if the resonance unit (RU) ratio varies between 0.5 and 2. Four compounds (red) were discarded by after visual analysis of the SPR signal. **B)** Structure of four primary hits with observable dose-dependent binding. The quality of the titration curve (upper panel) obtained by steady-state kinetic fitting of SPR sensorgrams (lower panel) permits the determination of a reliable K_D for compound #80, but only an approximation of K_D values for the other compounds. SPR sensorgrams were obtained at five concentrations (2.5 µM, black; 7.4 µM, tan; 22.2 µM, redbrick; 66.7 µM, greenblue; 200 µM, blue). Common pharmacophoric features (indole, phenol) are highlighted in orange and green, respectively.

The specifically-dedicated FitDock algorithm[47] fits the initial conformation of the hit to the given templates using a hierarchical multi-feature alignment approach, subsequently explores the possible conformations, and finally outputs a refined docking pose (**Figure 4**).

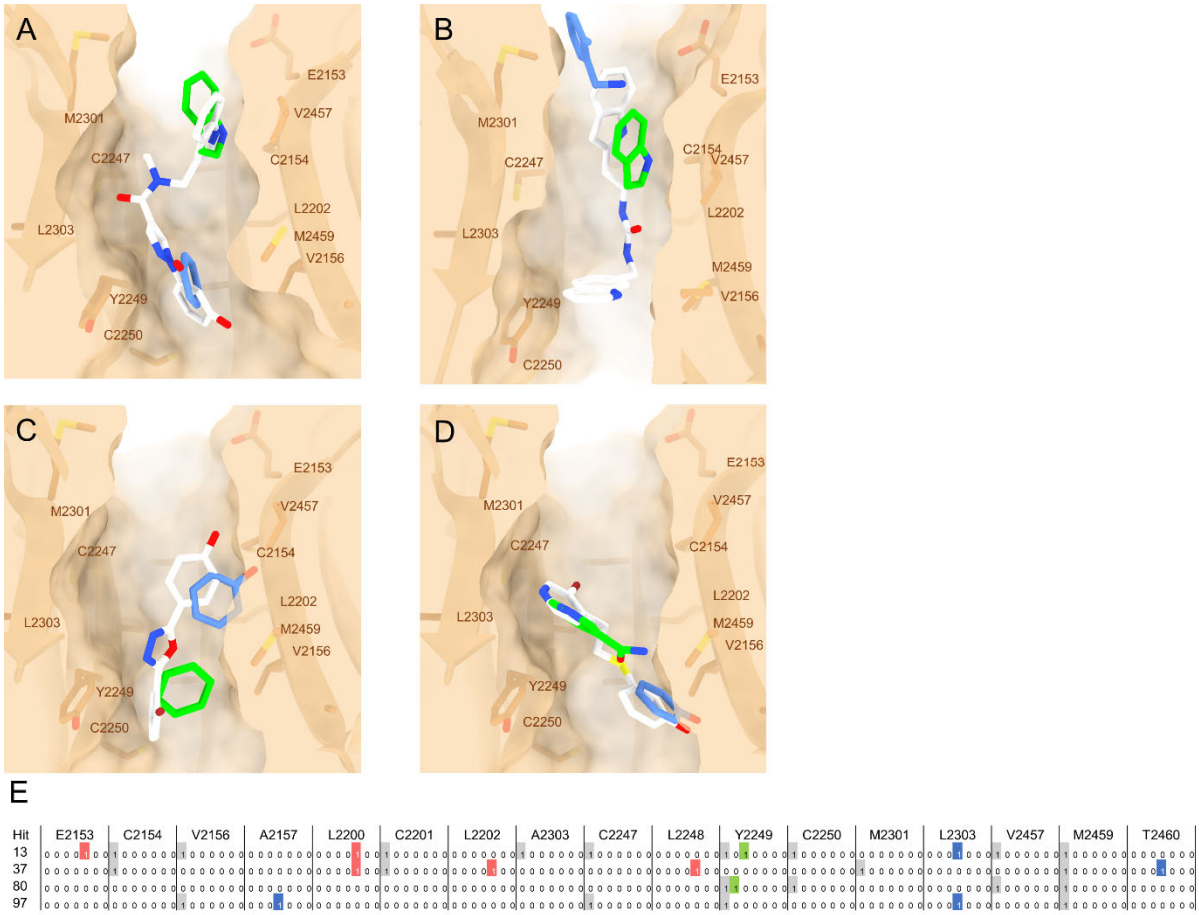

Figure 4. Hypothetical binding mode of Round 1 hits. Procare pose of selected fragments pairs (green, blue) to generate POEM hits (white) #13 **(A)**, #37 **(B)**, #80 **(C)** and #97 **(D)**. Interaction fingerprint of POEM hits **(E)**. For each LRRK2-WDR residue, non-covalent interactions are registered for every hit and outputted as a bit string (1st bit: hydrophobic, grey; 2nd bit: face-to-face aromatic, green; 3rd bit: edge-to-face aromatic, green; 4th bit: hydrogen-bond with protein donor, blue, 5th bit: hydrogen-bond with protein acceptor, red; 6th bit: ionic bond with protein positive ionized, dark blue; 7th bit: ionic bond with protein negative ionized, dark red).

Except for compound #37, the proposed docking poses were in line with the ProCare alignment of the corresponding seed fragments (**Figure 4**). Computing protein-ligand interaction fingerprints suggest some shared interactions (hydrogen bonds to main chain atoms of Leu2000, Leu2202 and Leu2303, aromatic interactions with Tyr2249), that should be considered with care since unconstrained docking with other tools (e.g. PLANTS, GOLD) did not either recapitulate any of

these docking poses nor agreed with the common consensus poses. In the absence of protein-ligand experimental structural data, the weak affinity of these binders and the cylindrical shape of the WDR pocket remain true hurdles to the precise identification of their binding mode. To further validate the candidate hits and their chemical series, a round 2 hit expansion was therefore undertaken by selecting and testing close analogs in REAL space.

Hit Expansion (Round 2). The second round of the challenge consisted in selecting chemically close analogs of round 1 hits (**Supporting Table S3**) to establish early structure-activity relationships (SAR) and fully confirm the primary hit as well as the corresponding chemotype. To this end, all hits were resynthesized therefore providing two independent batches of slightly different purity (>90%) for direct comparison. One hit (#37) could be confirmed with a comparable dissociation constant for the two batches (**Table 1**), and confirmed in an orthogonal differential scanning fluorimetry (DSF) assay (increase of the LRRK2 melting temperature by +0.7 °C at 100 µM, data not shown).

Table 1. Round 2 analogs tested

Hit	K_D, µM (% binding) Batch 1	K_D, µM (% binding) Batch 2	Observation	Ordered analogs	Synthesized analogs	Active analogs
13	37 (46%)	>200	Lower solubility and slow dissociation of batch 2	14	12	4
37	23 (27%)	37 (31%)	Confirmed hit	4	3	0
80	45 (99%)	220 (136%)	Lower solubility of batch 2	17	0	0
97	36 (37%)	>200	Lower solubility of batch 1	7	6	0

The three remaining compounds (#13, #80, and #97) exhibited discrepancies attributed to variations in solubilities and aggregation potential measured by dynamic light scattering (**Supporting Table S4**). For compounds #13 and #80, the lower solubility of the second batch precluded for reproducing the K_D value obtained by the more soluble first batch. Conversely, the higher solubility of the second batch (compound #97) invalidated the binding data associated with the less soluble initial batch (**Table 1**). Hit analogs were searched in the full REAL space of 36 billion synthesizable compounds by maximum common substructure (MCS) similarity searches, thanks to the SpaceMACS algorithm,[50] considering only 45 new compounds with a MCS similarity higher than 0.80 to their parent hit to ascertain chemotype conservation (**Supporting Table S3).**

Four analogs of hit #13 (#1202_07, #1202_10, #1202_11, and #1202_13; **Figure 5**) exhibited clear binding to LRRK2-WDR by SPR, albeit with lower K_D values (**Figure 5, Supporting Figure S3**).

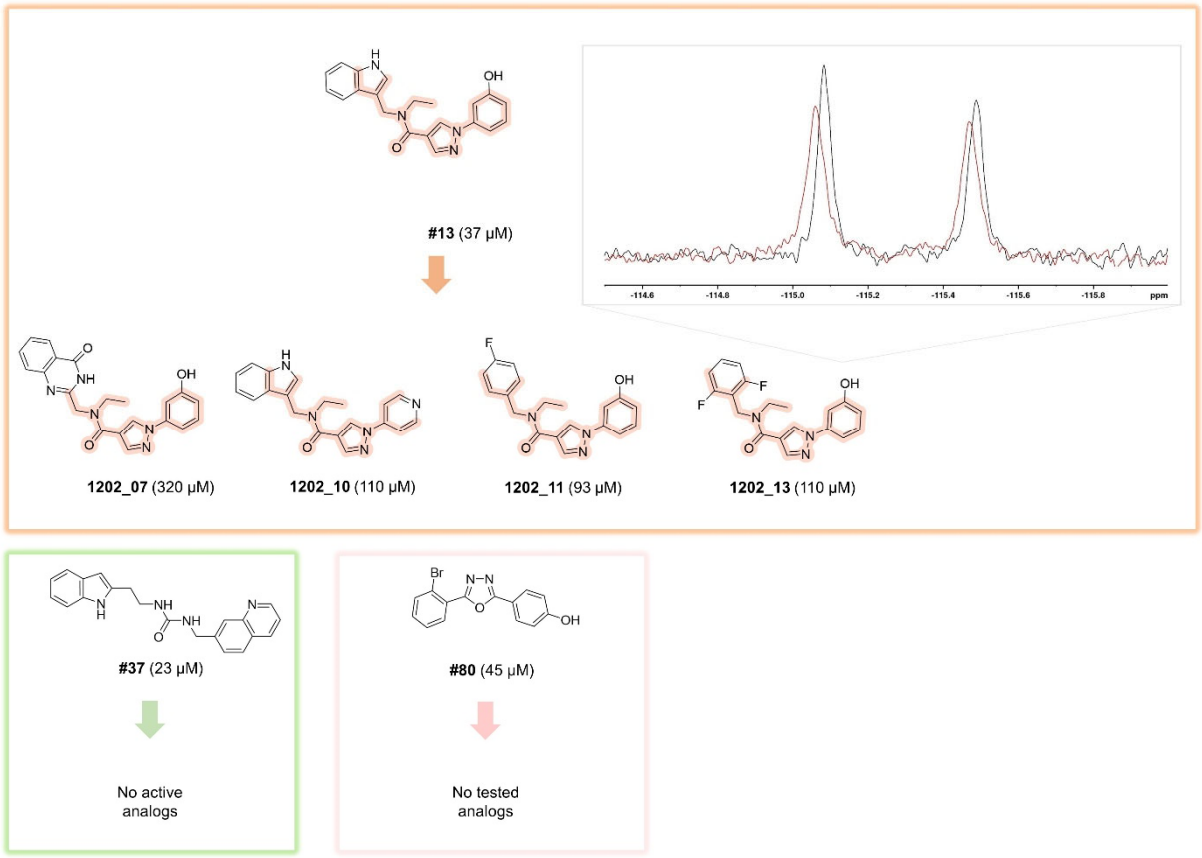

Figure 5. Round 2 hit expansion. Hit analogs for which a K_D could be determined by SPR, at a concentration where the compound is soluble, are indicated. Two primary hits could not be expanded either by lack of active analogs (#37) or the failure to synthesize compound derivatives (#80). The maximum common substructure between primary hits and their analogs is highlighted. In insert is shown the 1D-^{19}F spectrum of analog #1202_13 (10 mM, black), overlayed with its

spectrum upon mixing with 20 mM LRRK2-WDR (red), showing the downfield shift and broadening of its peaks due to binding.

The indole ring could be replaced by either fluorinated phenyls or the 4-quinazoline-one present in the hit #45, but with a markedly decreased affinity to LRRK2-WDR. The m-hydroxphenyl pharmacophoric feature seams of high importance since its replacement led to inactive analogs, with the exception of the 4-pyridinyl analog #1202_10. Interestingly, compound #1202_13, though binding weakly as measured by SPR (**Figure 5, Supporting Figure S3**), was confirmed to bind to the untagged LRRK2-WDR domain by ^{19}F-NMR, a very sensitive biophysical method for low affinity ligand detection[55] (**Figure 5**), thereby confirming the relevance of the corresponding chemotype in an orthogonal assay.

All three analogs of hit #37 that could be synthesized were found inactive, probably because of the absence of the crucial indole ring and the significant shortening of the linker. One very close analog of hit #45, lacking the 2-methyl group on the quinazolinone moiety (#HO_1202_22; **Supporting Table S3)** exhibited detectable but very weak binding to the target (data not shown). All other analogs (#HO_1202_20, #HO_1202_21, #HO_1202_23 and #HO_1202_24; **Supporting Table S3**) were found inactive despite minor changes at the quinazolinone ring, suggesting that the later moiety defines a potential pharmacophoric feature for WDR binding-ligands.

Unfortunately, none of 17 analogs of the promising hit #80 (K_D = 45 µM, 99% binding; **Figure 3**) could be evaluated although all these 2-5-diaryl 1,3,4-oxadiazoles are registered in the Enamine REAL Space. Last, none of the close analogs of hit #97 (**Supporting Table S3**) exhibited detectable binding in SPR experiments. Since the binding of the parent hit could not be reproduced after resynthesis of the compound, the corresponding series was not investigated further.

CONCLUSIONS

We present a structure-based approach (POEM) to early hit identification for orphan targets that is fully orthogonal to existing computational approaches (e.g. docking, pharmacophore search) and therefore expected to deliver complimentary information and hits. A clear advantage of the POEM workflow is the fuzziness in the initial subpocket to cavity similarity search that permits the location of preliminary poses of a few seed fragments, independent of any target-fragment interaction assessment, that would probably be missed by standard molecular docking. However, the deep generative linking of fragment hits remains the Achilles' heel of this method since target-dependent information is lost upon converting 3D seed fragment poses to final candidates which are enumerated as SMILES strings. Moreover, as any other generative algorithm not driven by existing organic chemistry reactions, final candidates are usually not directly purchasable among commercially available compound libraries or on-demand chemical spaces, forcing the user to select the closest possible analogs for which a common binding mode is expected.

Despite the above-mentioned limitations, and when applied to an *a priori* difficult cavity (LRRK2-WDR), the method was able to quickly provide four experimentally validated hits of potencies comparable to that reached with much more computer-intensive approaches (e.g. docking of ultra-large chemical spaces) and with comparable hit rates (ca. 5%).[56-57] Unfortunately, none of the primary hits could be further optimized for potency in a preliminary follow-up structure-activity relationship (SAR) by catalog.[58]

Interestingly, the CACHE challenge #1 illustrates the difficulty in interpreting SAR data when structure-activity and structure-solubility relationships of weakly potent hits are deeply interdependent. Resynthesizing different batches for every hit and interpreting their potency in the light of solubility and aggregation data remains good practice to avoid propagating false positives and to select the most valuable hits for further optimization.

ACKNOWLEDGMENTS

We deeply thank the cc-IN2P3 calculation center (Villeurbanne, France) for allocation of computing resources and excellent support. We thank Zarah Hejazi and Carla Kharadjian, Maria Kutera and Helen Li for their contribution to protein production.

FUNDING

The authors are thankful to the doctoral school of chemical sciences (EDSC, University of Strasbourg) for a Ph.D. grant to M.E. The CACHE #1 challenge was supported by Conscience thanks to funding from the Canadian Strategic Innovation Fund (SIF Stream 5) and by the Michael J. Fox foundation. The Structural Genomics Consortium is a registered charity (no: 1097737) that receives funds from Bayer AG, Boehringer Ingelheim, Bristol Myers Squibb, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking [EUbOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer and Takeda. NMR assays were carried out on spectrometers generously supported by the Princess Margaret Cancer Foundation.

Supporting Information.

Filtering rules to select drug-like compounds, list of 94 virtual hits selected for experimental validation, round 2 analogues proposed for synthesis and experimental validation, topological requirements to connect fragment atoms by a linker, structures of original POEM hits leading to experimentally validated LRRK2-WDR binders, Characterization of LRRK2 ligand 1202_13.

Data and Software Availability

Data. All data generated by the CACHE #1 challenge have been made public and are available at https://cache-challenge.org/results-cache-challenge-1 including: an excel file with the annotated ranking of the computational methods based on the experimental data, a schematic summary of the top performing computational methods, a slide deck with detailed Round 2 data for each participant, an excel file with Round 1 hits that were advanced to Round 2, an excel file with all data for all compounds for both rounds, short manuals for non-experts to understand SPR and DLS data.

The sc-PDB (v.2022) database of druggable protein-ligand complexes can be downloaded from https://seafile.unistra.fr/library/18e57c6f-d2af-46ac-a4fb-f2bc5008e66c/sc-PDB/2022. The REAL diversity set is provided by Enamine (https://enamine.net/compound-collections/real-compounds/real-database-subsets).

Software. ProCare (version 0.1.2) is available at https://github.com/kimeguida/ProCare. IChem (version 5.2.9) was downloaded from http://bioinfo-pharma.u-strasbg.fr/labwebsite/download.html. DeLinker was retrieved from https://github.com/oxpig/DeLinker. OpenEye's Filter (version 4.2.1.1) is available from Cadence Molecular Science (https://www.eyesopen.com). RDKit (version 2019.03.4.0) was downloaded from https://www.rdkit.org/. Pipeline Pilot (version 22.1.0.2935) is available from Dassault Systèmes Biovia Corp. (https://www.3ds.com/products/biovia/pipeline-pilot). Corina (version 3.40) was obtained from Molecular Networks GmbH (https://mn-am.com/products/corina/). FitDock was downloaded from http://cao.labshare.cn. Open Babel (version 3.1.0) can be retrieved from https://github.com/openbabel/openbabel/releases. SpaceMACS (version 0.9.2) is available from BiosolveIT (https://www.biosolveit.de/products/#SpaceMACS).

REFERENCES

1. Bender, B. J.; Gahbauer, S.; Luttens, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balius, T. E.; Carlsson, J.; Irwin, J. J.; Shoichet, B. K., A Practical Guide to Large-Scale Docking. *Nat Protoc,* **2021,** *16*, 4799-4832.
2. Eguida, M.; Rognan, D., Estimating the Similarity between Protein Pockets. *Int J Mol Sci,* **2022,** *23*.
3. Carpenter, K. A.; Altman, R. B., Databases of Ligand-Binding Pockets and Protein-Ligand Interactions. *Comput Struct Biotechnol J,* **2024,** *23*, 1320-1338.
4. Di Palma, F.; Abate, C.; Decherchi, S.; Cavalli, A., Ligandability and Druggability Assessment Via Machine Learning. *Wires Comput Mol Sci,* **2023,** *13*.
5. Ehrt, C.; Brinkjost, T.; Koch, O., Impact of Binding Site Comparisons on Medicinal Chemistry and Rational Molecular Design. *J Med Chem,* **2016,** *59*, 4121-4151.
6. Melancon, K.; Pliushcheuskaya, P.; Meiler, J.; Kunze, G., Targeting Ion Channels with Ultra-Large Library Screening for Hit Discovery. *Front Mol Neurosci,* **2023,** *16*, 1336004.
7. Trisciuzzi, D.; Villoutreix, B. O.; Siragusa, L.; Baroni, M.; Cruciani, G.; Nicolotti, O., Targeting Protein-Protein Interactions with Low Molecular Weight and Short Peptide Modulators: Insights on Disease Pathways and Starting Points for Drug Discovery. *Expert Opin Drug Discov,* **2023,** *18*, 737-752.
8. Schmitt, S.; Kuhn, D.; Klebe, G., A New Method to Detect Related Function among Proteins Independent of Sequence and Fold Homology. *J Mol Biol,* **2002,** *323*, 387-406.
9. De Franchi, E.; Schalon, C.; Messa, M.; Onofri, F.; Benfenati, F.; Rognan, D., Binding of Protein Kinase Inhibitors to Synapsin I Inferred from Pair-Wise Binding Site Similarity Measurements. *Plos One,* **2010,** *5*.
10. Eguida, M.; Rognan, D., A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. *J Med Chem,* **2020,** *63*, 7127-7142.

11. Eguida, M.; Rognan, D., Unexpected Similarity between Hiv-1 Reverse Transcriptase and Tumor Necrosis Factor Binding Sites Revealed by Computer Vision. *J Cheminform,* **2021,** *13*, 90.

12. Eguida, M.; Schmitt-Valencia, C.; Hibert, M.; Villa, P.; Rognan, D., Target-Focused Library Design by Pocket-Applied Computer Vision and Fragment Deep Generative Linking. *J Med Chem,* **2022,** *65*, 13771-13783.

13. Stanley, M.; Segler, M., Fake It until You Make It? Generative De Novo Design and Virtual Screening of Synthesizable Molecules. *Curr Opin Struct Biol,* **2023,** *82*, 102658.

14. Schneider, G.; Fechner, U., Computer-Based De Novo Design of Drug-Like Molecules. *Nat Rev Drug Discov,* **2005,** *4*, 649-663.

15. Schneider, G.; Fechner, U., Computer-Based Design of Drug-Like Molecules. *Nat Rev Drug Discov,* **2005,** *4*, 649-663.

16. Pierce, A. C.; Rao, G.; Bemis, G. W., BREED: Generating Novel Inhibitors through Hybridization of Known Ligands. Application to Cdk2, P38, and Hiv Protease. *J Med Chem,* **2004,** *47*, 2768-2775.

17. Sydow, D.; Schmiel, P.; Mortier, J.; Volkamer, A., KinFragLib: Exploring the Kinase Inhibitor Space Using Subpocket-Focused Fragmentation and Recombination. *J Chem Inf Model,* **2020,** *60*, 6081-6094.

18. Lewis, R. A.; Roe, D. C.; Huang, C.; Ferrin, T. E.; Langridge, R.; Kuntz, I. D., Automated Site-Directed Drug Design Using Molecular Lattices. *J Mol Graph,* **1992,** *10*, 66-78, 106.

19. Pearlman, D. A.; Murcko, M. A., Concepts - New Dynamic Algorithm for De-Novo Drug Suggestion. *J Comput Chem,* **1993,** *14*, 1184-1193.

20. Ramensky, V.; Sobol, A.; Zaitseva, N.; Rubinov, A.; Zosimov, V., A Novel Approach to Local Similarity of Protein Binding Sites Substantially Improves Computational Drug Design Results. *Proteins,* **2007,** *69*, 349-357.

21. Bohm, H. J., The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors. *J Comput Aided Mol Des,* **1992,** *6*, 61-78.

22. Wang, R. X.; Gao, Y.; Lai, L. H., LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *J Mol Model,* **2000,** *6*, 498-516.

23. Penner, P.; Martiny, V.; Gohier, A.; Gastreich, M.; Ducrot, P.; Brown, D.; Rarey, M., Shape-Based Descriptors for Efficient Structure-Based Fragment Growing. *J Chem Inf Model,* **2020,** *60*, 6269-6281.

24. Rotstein, S. H.; Murcko, M. A., Groupbuild: A Fragment-Based Method for De Novo Drug Design. *J Med Chem,* **1993,** *36*, 1700-1710.

25. Lewis, R. A., Automated Site-Directed Drug Design: A Method for the Generation of General Three-Dimensional Molecular Graphs. *J Mol Graph,* **1992,** *10*, 131-143.

26. Moriaud, F.; Doppelt-Azeroual, O.; Martin, L.; Oguievetskaia, K.; Koch, K.; Vorotyntsev, A.; Adcock, S. A.; Delfaud, F., Computational Fragment-Based Approach at PDB Scale by Protein Local Similarity. *J Chem Inf Model,* **2009,** *49*, 280-294.

27. Durrant, J. D.; Friedman, A. J.; McCammon, J. A., CrystalDock: A Novel Approach to Fragment-Based Drug Design. *J Chem Inf Model,* **2011,** *51*, 2573-2580.

28. Ackloo, S.; Al-Awar, R.; Amaro, R. E.; Arrowsmith, C. H.; Azevedo, H.; Batey, R. A.; Bengio, Y.; Betz, U. A. K.; Bologa, C. G.; Chodera, J. D.; Cornell, W. D.; Dunham, I.; Ecker, G. F.; Edfeldt, K.; Edwards, A. M.; Gilson, M. K.; Gordijo, C. R.; Hessler, G.; Hillisch, A.; Hogner, A.; Irwin, J. J.; Jansen, J. M.; Kuhn, D.; Leach, A. R.; Lee, A. A.; Lessel, U.; Morgan, M. R.; Moult, J.; Muegge, I.; Oprea, T. I.; Perry, B. G.; Riley, P.; Rousseaux, S. A. L.; Saikatendu, K. S.; Santhakumar, V.; Schapira, M.; Scholten, C.; Todd, M. H.; Vedadi, M.; Volkamer, A.; Willson, T. M., Cache (Critical Assessment of Computational Hit-Finding Experiments): A Public-Private Partnership Benchmarking Initiative to Enable the Development of Computational Methods for Hit-Finding. *Nat Rev Chem,* **2022,** *6*, 287-295.

29. Myasnikov, A.; Zhu, H. W.; Hixson, P.; Xie, B. E.; Yu, K. W.; Pitre, A.; Peng, J. M.; Sun, J., Structural Analysis of the Full-Length Human LRRK2. *Cell,* **2021,** *184*, 3519-3527.

30. Deniston, C. K.; Salogiannis, J.; Mathea, S.; Snead, D. M.; Lahiri, I.; Matyszewski, M.; Donosa, O.; Watanabe, R.; Bohning, J.; Shiau, A. K.; Knapp, S.; Villa, E.; Reck-Peterson, S. L.; Leschziner, A. E., Structure of LRRK2 in Parkinson's Disease and Model for Microtubule Interaction. *Nature,* **2020,** *588*, 344-349.

31. Tolosa, E.; Vila, M.; Klein, C.; Rascol, O., LRRK2 in Parkinson Disease: Challenges of Clinical Trials. *Nat Rev Neurol,* **2020,** *16*, 97-107.

32. Wojewska, D. N.; Kortholt, A., LRRK2 Targeting Strategies as Potential Treatment of Parkinson's Disease. *Biomolecules,* **2021,** *11*.

33. Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M., Protoss: A Holistic Approach to Predict Tautomers and Protonation States in Protein-Ligand Complexes. *J Cheminform,* **2014,** *6*, 12.

34. Da Silva, F.; Desaphy, J.; Rognan, D., IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein-Ligand Interactions. *ChemMedChem,* **2018,** *13*, 507-510.

35. Desaphy, J.; Azdimousa, K.; Kellenberger, E.; Rognan, D., Comparison and Druggability Prediction of Protein-Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. *J Chem Inf Model,* **2012,** *52*, 2287-2299.

36. Desaphy, J.; Rognan, D., sc-PDB-Frag: A Database of Protein-Ligand Interaction Patterns for Bioisosteric Replacements. *J Chem Inf Model,* **2014,** *54*, 1908-1918.

37. Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M., Recap--Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J Chem Inf Comput Sci,* **1998,** *38*, 511-522.

38. Certara USA, Inc., Princeton, NJ 08540, U.S.A.

39. Zhang, P.; Fan, Y.; Ru, H.; Wang, L.; Magupalli, V. G.; Taylor, S. S.; Alessi, D. R.; Wu, H., Crystal Structure of the Wd40 Domain Dimer of LRRK2. *Proc Natl Acad Sci U S A,* **2019,** *116*, 1579-1584.

40. https://www.rcsb.org/structure/6dlo (accessed 04-20-2024).

41. Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M., Deep Generative Models for 3D Linker Design. *J Chem Inf Model,* **2020,** *60*, 1983-1995.

42. OpenEye Scientific Software, Santa Fe, NM 87508, U.S.A. (accessed 04-20-2022).

43. Real Diversity Set, https://enamine.net/compound-collections/real-compounds/real-compound-libraries (accessed 04-20-2024).

44. RdKit: Open-Source Cheminformatics Software, http://www.rdkit.org (accessed 04-20-2024).

45. Dassault Systèmes Biovia Corp, San Diego, CA.

46. Molecular Networks GmbH, Nürnberg, Germany. https://mn-am.com/products/corina/ (accessed 03-03-2024).

47. Yang, X. C.; Liu, Y.; Gan, J. H.; Xiao, Z. X.; Cao, Y., FitDock: Protein-Ligand Docking by Template Fitting. *Brief Bioinform,* **2022,** *23*.

48. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open Chemical Toolbox. *J Cheminform,* **2011,** *3*, 33.

49. Readily-Accessible on-Demand Chemical Spaces, https://www.biosolveit.de/infinisee (accessed 04-20-2024).

50. Schmidt, R.; Klein, R.; Rarey, M., Maximum Common Substructure Searching in Combinatorial Make-on-Demand Compound Spaces. *J Chem Inf Model,* **2022,** *62*, 2133-2150.

51. Hutchinson, A.; Seitova, A., Production of Recombinant PRMT Proteins Using the Baculovirus Expression Vector System. *J Vis Exp,* **2021**.

52. Allali-Hassani, A.; Szewczyk, M. M.; Ivanochko, D.; Organ, S. L.; Bok, J.; Ho, J. S. Y.; Gay, F. P. H.; Li, F.; Blazer, L.; Eram, M. S.; Halabelian, L.; Dilworth, D.; Luciani, G. M.; Lima-Fernandes,

E.; Wu, Q.; Loppnau, P.; Palmer, N.; Talib, S. Z. A.; Brown, P. J.; Schapira, M.; Kaldis, P.; O'Hagan, R. C.; Guccione, E.; Barsyte-Lovejoy, D.; Arrowsmith, C. H.; Sanders, J. M.; Kattar, S. D.; Bennett, D. J.; Nicholson, B.; Vedadi, M., Discovery of a Chemical Probe for PRDM9. *Nature Communications,* **2019,** *10*, 5759.

53. Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E., sc-PDB: A 3D-Database of Ligandable Binding Sites--10 Years On. *Nucleic Acids Res,* **2015,** *43*, D399-404.

54. Ertl, P.; Schuffenhauer, A., Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. *J Cheminform,* **2009,** *1*, 8.

55. Dalvit, C.; Vulpetti, A., Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J Med Chem,* **2019,** *62*, 2218-2244.

56. https://cache-challenge.org/results-cache-challenge-1 (accessed 04-20-2024).

57. Kaplan, A. L.; Confair, D. N.; Kim, K.; Barros-Alvarez, X.; Rodriguiz, R. M.; Yang, Y.; Kweon, O. S.; Che, T.; McCorvy, J. D.; Kamber, D. N.; Phelan, J. P.; Martins, L. C.; Pogorelov, V. M.; DiBerto, J. F.; Slocum, S. T.; Huang, X. P.; Kumar, J. M.; Robertson, M. J.; Panova, O.; Seven, A. B.; Wetsel, A. Q.; Wetsel, W. C.; Irwin, J. J.; Skiniotis, G.; Shoichet, B. K.; Roth, B. L.; Ellman, J. A., Bespoke Library Docking for 5-HT(2a) Receptor Agonists with Antidepressant Activity. *Nature,* **2022,** *610*, 582-591.

58. Klingler, F. M.; Gastreich, M.; Grygorenko, O. O.; Savych, O.; Borysko, P.; Griniukova, A.; Gubina, K. E.; Lemmen, C.; Moroz, Y. S., SAR by Space: Enriching Hit Sets from the Chemical Space. *Molecules,* **2019,** *24*.

Table of content graphics