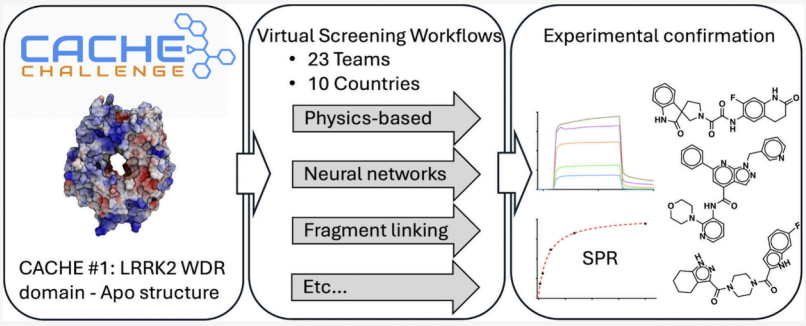
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

CACHE Challenge #1: Targeting the WDR Domain of LRRK2, A Parkinson's Disease Associated Protein

Fengling Li, Suzanne Ackloo, Cheryl H. Arrowsmith, Fuqiang Ban, Christopher J. Barden, Hartmut Beck, Jan Beránek, Francois Berenger, Albina Bolotokova, Guillaume Bret, Marko Breznik, Emanuele Carosati, Irene Chau, Yu Chen, Artem Cherkasov, Dennis Della Corte, Katrin Denzinger, Aiping Dong, Sorin Draga, Ian Dunn, Kristina Edfeldt, Aled Edwards, Merveille Eguida, Paul Eisenhuth, Lukas Friedrich, Alexander Fuerll, Spencer S Gardiner, Francesco Gentile, Pegah Ghiabi, Elisa Gibson, Marta Glavatskikh, Christoph Gorgulla, Judith Guenther, Anders Gunnarsson, Filipp Gusev, Evgeny Gutkin, Levon Halabelian, Rachel J. Harding, Alexander Hillisch, Laurent Hoffer, Anders Hogner, Scott Houliston, John J Irwin, Olexandr Isayev, Aleksandra Ivanova, Celien Jacquemard, Austin J Jarrett, Jan H. Jensen, Dmitri Kireev, Julian Kleber, S. Benjamin Koby, David Koes, Ashutosh Kumar, Maria G. Kurnikova, Alina Kutlushina, Uta Lessel, Fabian Liessmann, Sijie Liu, Wei Lu, Jens Meiler, Akhila Mettu, Guzel Minibaeva, Rocco Moretti, Connor J Morris, Chamali Narangoda, Theresa Noonan, Leon Obendorf, Szymon Pach, Amit Pandit, Sumera Perveen, Gennady Poda, Pavel Polishchuk, Kristina Puls, Vera Pütter, Didier Rognan, Dylan Roskams-Edris, Christina Schindler, François Sindt, Vojtěch Spiwok, Casper Steinmann, Rick L. Stevens, Valerij Talagayev, Damon Tingey, Oanh Vu, W. Patrick Walters, Xiaowen Wang, Zhenyu Wang, Gerhard Wolber, Clemens Alexander Wolf, Lars Wortmann, Hong Zeng, Carlos A. Zepeda, Kam Y. J. Zhang, Jixian Zhang, Shuangjia Zheng, and Matthieu Schapira*

Cite This: https://doi.org/10.1021/acs.jcim.4c01267 Read Online



ABSTRACT: The CACHE challenges are a series of prospective benchmarking exercises to evaluate progress in the field of computational hit-finding. Here we report the results of the inaugural CACHE challenge in which 23 computational teams each selected up to 100 commercially available compounds that they predicted would bind to the WDR domain of the Parkinson's disease target LRRK2, a domain with no known ligand and only an apo structure in the PDB. The lack of known binding data and presumably low druggability of the target is a challenge to computational hit finding methods. Of the 1955 molecules predicted by participants in Round 1 of the challenge, 73 were found to bind to LRRK2 in an SPR assay with a K_D lower than 150 μM. These 73

continued...

Received: July 18, 2024
Revised: October 21, 2024
Accepted: October 28, 2024

molecules were advanced to the Round 2 hit expansion phase, where computational teams each selected up to 50 analogs. Binding was observed in two orthogonal assays for seven chemically diverse series, with affinities ranging from 18 to 140 μM. The seven successful computational workflows varied in their screening strategies and techniques. Three used molecular dynamics to produce a conformational ensemble of the targeted site, three included a fragment docking step, three implemented a generative design strategy and five used one or more deep learning steps. CACHE #1 reflects a highly exploratory phase in computational drug design where participants adopted strikingly diverging screening strategies. Machine learning-accelerated methods achieved similar results to brute force (e.g., exhaustive) docking. First-in-class, experimentally confirmed compounds were rare and weakly potent, indicating that recent advances are not sufficient to effectively address challenging targets.

INTRODUCTION

The Critical Assessment of Computational Hit-finding Experiments (CACHE) Challenges are a triannual series of prospective benchmarking exercises. In the first round of each challenge, computational chemistry experts are invited to select up to 100 compounds from commercial libraries that they predict bind to a predefined target. Compounds are purchased and binding to the target protein is tested experimentally. Compounds of interest are then advanced to Round 2, a hit expansion round where participants select up to 50 follow-up molecules for experimental testing. Based on both rounds, an independent committee composed of industry experts assesses the validity of the biophysical activity data of each series, the drug-likeness of the validated hits, and their suitability as starting points for hit-to-lead optimization. Both the structures and bioactivity data serve to identify the best-performing computational methods, after which all data are publicly released on https://cache-challenge.org/. The goal of CACHE is to provide an objective and transparent forum where a diverse array of virtual screening workflows are compared against the same protein target and evaluated using the same experimental assays and platform.[1] Unlike other benchmarking challenges such as CSAR, D3R, SAMPL or CELPP, CACHE challenges are prospective in that predictions are made before experimental data are generated.[2−5]

The first CACHE challenge focused on leucine-rich repeat kinase 2 (LRRK2), the most mutated protein in familial Parkinson's disease (PD). Mutations in the kinase domain of LRRK2 can increase its activity, leading to pathogenic hallmarks associated with PD.[6−8] While LRRK2 kinase activity has been an active area of drug discovery, the first-generation LRRK2 kinase inhibitors have not shown the expected therapeutic benefit. This may be due to LRRK2's scaffolding function[9] or the distinct conformational states stabilized by Type I and Type II inhibitors.[10]

An alternative and overlooked strategy to inhibit pathogenic LRRK2 is to pharmacologically target its WD40 repeat (WDR) domain (LRRK2-WDR), which is juxtaposed to the kinase domain[11] but has no clear function or known interactor (Figure 1). WDR domains are typically protein interaction hubs, a number of which have been linked to disease and have been identified as druggable targets.[12,13] Despite their canonical donut-like structure, residues lining the central cavity of WDR domains are not conserved, leading to high ligand selectivity. In the case of LRRK2, the WDR domain may be important for recruiting binding partners or for binding with tubulin.

The WDR domain is also relevant to PD pathogenesis. A disease-linked mutation in this domain located at the interface of the LRRK2-WDR dimer enhances LRRK2 kinase activity and antagonizes dimerization.[11] Identifying compounds that bind to the LRRK2-WDR domain presents a potentially novel approach to targeting this protein, though no ligand was reported to date. In this challenge, participants were tasked with using the apo structure of LRRK2-WDR (PDB code 6DLO)[11] to predict compounds that could occupy the central cavity of the donut-shaped domain (Figure 1).

We provide here an overview of the first CACHE challenge, where 23 research teams from ten countries collectively predicted 1955 compounds targeting LRRK2-WDR. After a hit identification round (Round 1) followed by hit expansion (Round 2), seven chemical series predicted by seven participants produced convincing binding data in two orthogonal assays. These compounds are the first reported that target LRRK2-WDR and represent valuable chemical starting points for hit-to-lead optimization. Computational workflows were diverse and often included a step driven by deep learning. Hit rates were low and most compounds bound with an affinity above 50 μM, reflecting the challenges of structure-based virtual screening when only an apo form of the targeted binding pocket and no ligand is available.

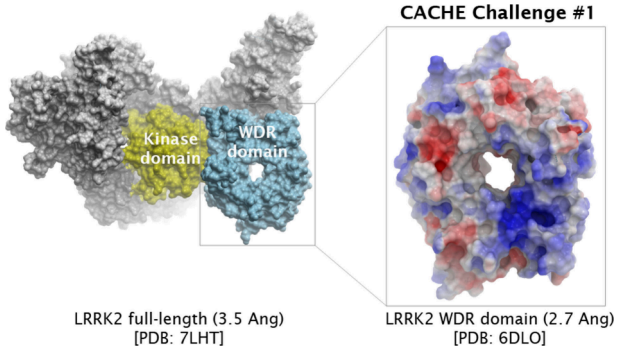


Figure 1. CACHE Challenge #1: Predicting ligands binding the central cavity of the LRRK2-WDR domain. Left: The kinase and WDR domains of LRRK2 are highlighted in the context of a LRRK2 monomer (PDB: 7LHT). Right: Electrostatic potential map of the LRRK2-WDR domain (blue, electropositive; red, electronegative).

RESULTS

Computational Workflows Were Diverse. Participating teams submitted applications and were selected based on a double-blind peer review process. Each team was asked to rate five applications, after which an independent Applications Review Committee (Table S1) undertook a final evaluation to verify the integrity of the peer review process. The 25 top-scoring teams from the double-blind peer review process were invited to participate (a cap dictated by experimental costs). Of those 25, 23 completed the challenge. Participants remained anonymous until the final release of the data at the end of the challenge, at which point they had the option to be deanonymized.

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

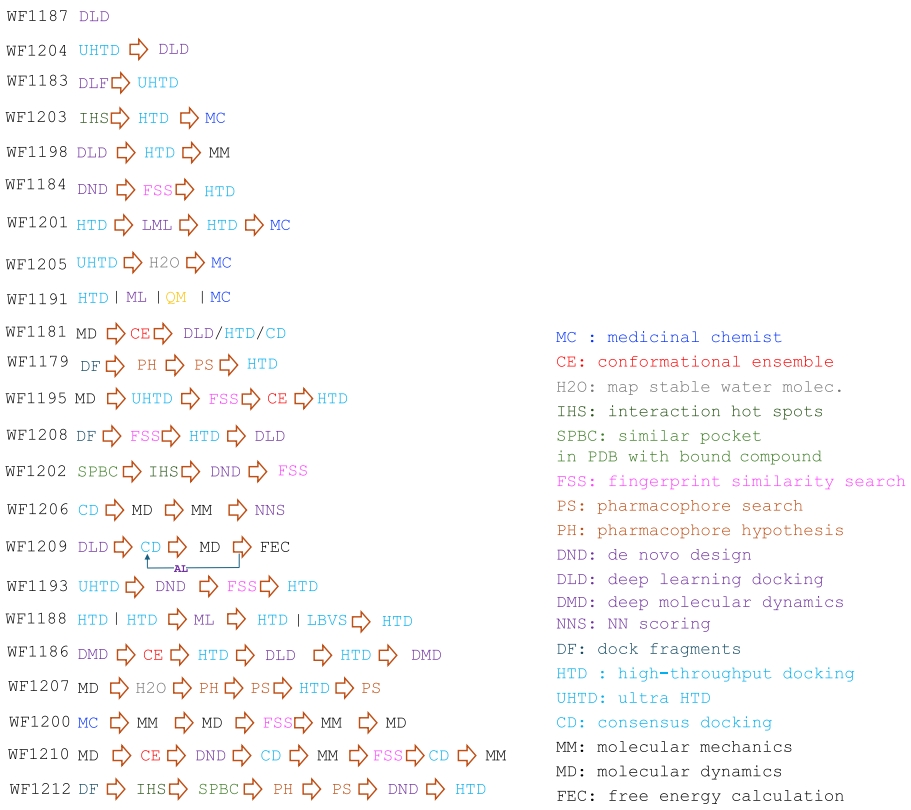


Figure 2. Computational cascades deployed in CACHE #1. Arrows denote cascading steps. "|" denotes alternative methods tested in parallel. Each workflow had a maximum credit of 100 compounds in Round 1, regardless of the number of methods tested.

CACHE #1 participants deployed a highly diverse set of computational tools and workflows, reflecting different hit selection strategies. The workflows are summarized in Figure 2 and described in detail at https://cache-challenge.org/challenges/predict-hits-for-the-wdr-domain-of-lrrk2/computational-methods. The following two examples demonstrate the significant divergence between two high-performing methods. In one instance, Shuangjia Zheng at Shanghai Jiao Tong University (workflow WF1187) used a multiscale and multitask neural network pretrained on ChEMBL and PubChem data as a one-step virtual screening workflow to produce the final compound selection, refined with physicochemical drug-likeness filters.[14] In contrast, Pavel Polishchuk at Palacky University (WF1210) adopted a screening cascade composed of seven distinct steps, where he first used molecular dynamics (MD) to generate a conformational ensemble of the binding pocket to which fragments were docked, grown and fine-tuned by a genetic algorithm for denovo ligand design; a consensus docking step refined with Molecular Mechanics generalized Born Surface Area (MM-GBSA) simulations was used to select the most promising ligands, commercial analogs of which (found by a fingerprint similarity search) were subjected to consensus docking followed by MM-GBSA to produce the final selection.

Between these two extremes, which both ranked in the top 10 after experimental testing, screening cascades varied significantly in the number and type of techniques deployed (Figure 2). Physics-based docking was used in 19 workflows; 12 incorporated at least one deep learning screening step, including deep learning docking in eight. Fragment-based approaches were adopted in five, four used MD to generate a conformational ensemble of the binding site, and four included consensus docking.

Selected Compounds Were Drug-like and Chemically Diverse. The 23 teams were given two months to conduct their virtual screens, after which they submitted a file of up to 100 compounds (or slightly more to account for failed synthesis) predicted to bind the central pocket of LRRK2-WDR and available from the Enamine REAL database (36 billion molecules at the time). Compounds had to satisfy three conditions: MW < 550 Da, cLogP < 5 and no reactive group. Participants were also encouraged to use badapple (https://datascience.unm.edu/badapple/)[15] to filter promiscuous compounds, though doing so was not mandatory. Almost all compounds (1875) were procured from Enamine, with a synthesis success rate of 93%. Another 80 were procured from MCULE. For most participants, combined procurements from these sources lead to a total of 80 to 100 compounds per team, with a few exceptions, including participant 1183 who selected only 37 compounds and participant 1188 for whom the success rate of chemical synthesis was higher than expected (113 compounds). The distribution of physicochemical descriptors of the 1955 compounds selected by the 23 participants reflected overall drug-like molecules (Figure 3, Table S2).

The compounds were chemically diverse, with 1629 out of 1955 having a Tanimoto distance greater than 0.3 with any compound selected by another participant (using 1536-bit fingerprints implemented in ICM, Molsoft LLC) (Figure 3c,d). The hit rate (% of compounds advancing to Round 2) was not higher for pairs of similar compounds (Tanimoto distance <0.3) than for the full set of 1955 molecules (3.6% and 3.7% respectively). Chemical diversity was also observed within selections from each participant, though some participants did

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

a

Participant	Compounds tested experimentally	Compounds advancing to Round 2	Hit rate
1179	100	2	2
1181	84	2	2
1183	37	2	5
1184	65	2	3
1186	99	11	11
1187	72	4	6
1188	113	3	3
1191	95	0	0
1193	92	4	4
1195	84	10	12
1198	79	0	0
1200	91	2	2
1201	101	1	1
1202	94	5	5
1203	105	2	2
1204	71	0	0
1205	98	3	3
1206	83	0	0
1207	101	1	1
1208	100	4	4
1209	59	7	12
1210	82	8	10
1212	50	0	0

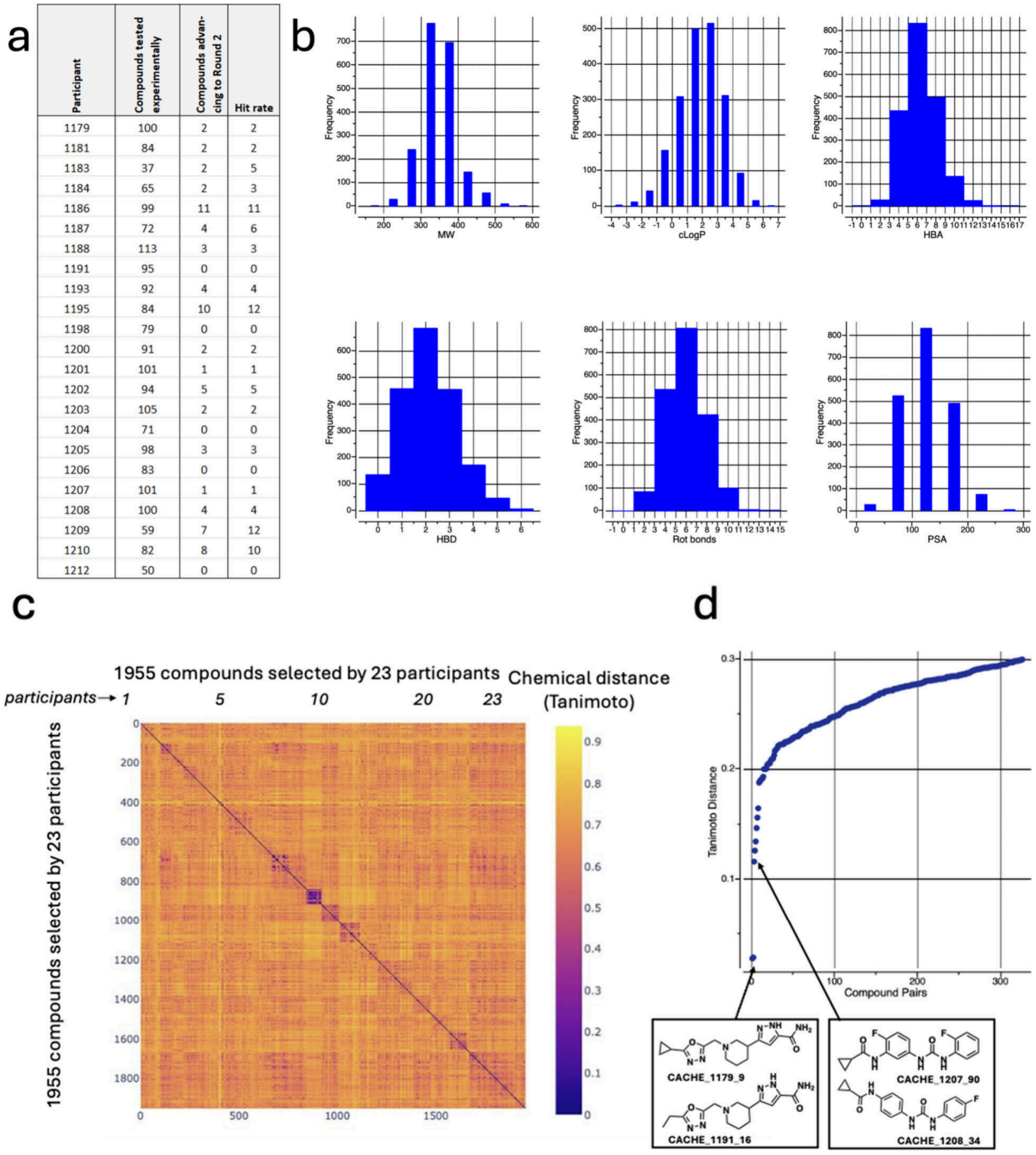


Figure 3. Drug-likeness and chemical diversity of selected compounds. (a) Number of compounds procured for each participant in the hit identification phase (Round 1) and number advancing to hit expansion (Round 2). (b) Molecular weight, calculated LogP, number of hydrogen bond acceptors/donors and rotatable bonds, and polar surface area distributions of the 1955 compounds. (c) Pairwise Tanimoto distance matrix of all pairs of compounds. (d) Tanimoto distance distribution of the 326 closest compound pairs (pairs of compounds selected by the same participant are not included).

select multiple chemically related compounds (dark squares along the diagonal in Figure 3c).

Experimental Testing of Round 1 Compounds. Binding of the 1955 Round 1 compounds to LRRK2-WDR was tested independently at 50 μM and 100 μM in a surface plasmon resonance (SPR) assay (Figure 4a, Table S3). 440 compounds with a R/R_{max} binding ratio (measured versus expected response unit (RU)) above 50% (i.e., significant binding) and below 200% (i.e., limited signs of nonspecific binding) in at least one of the two runs were evaluated in dose−response experiments. In

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

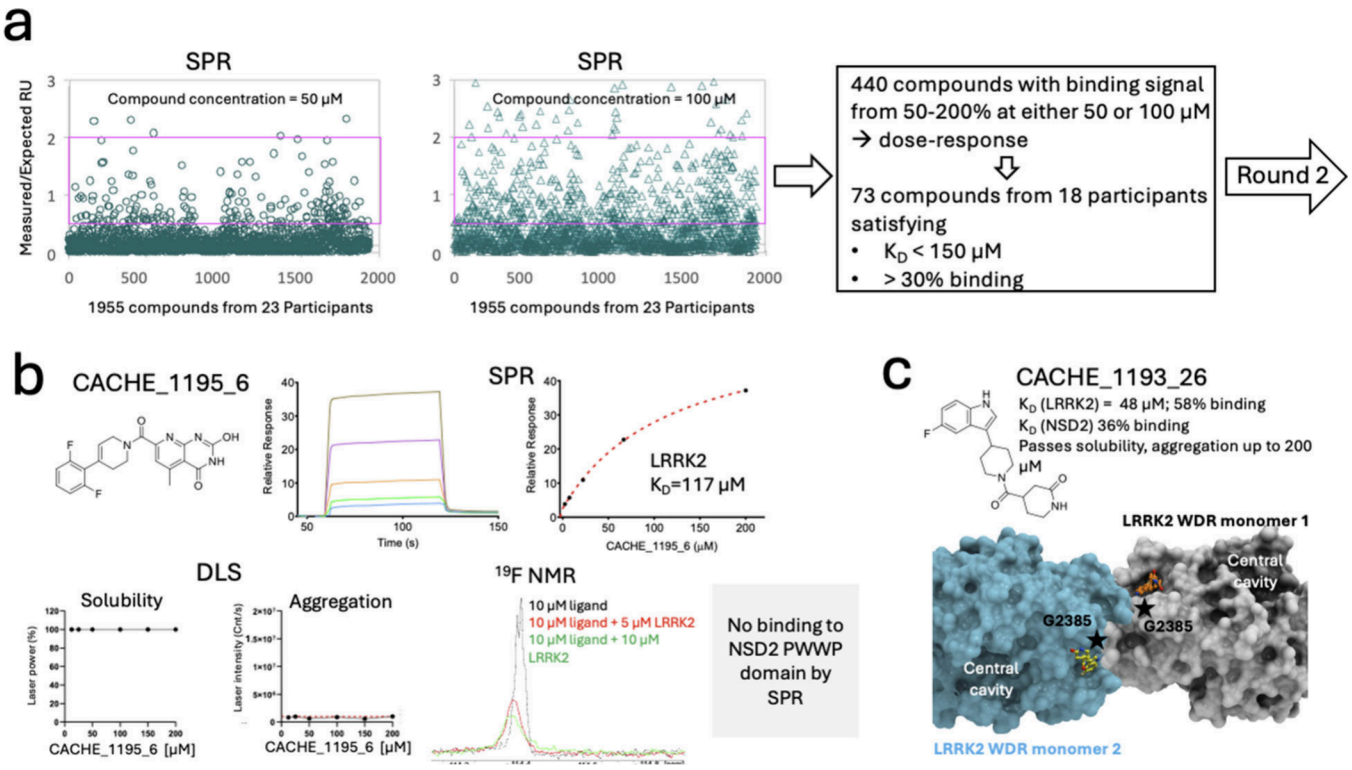


Figure 4. Experimental evaluation of CACHE #1 Round 1 compounds. (a) Binding to LRRK2 measured by SPR was used to advance compounds to Round 2. (b) Experimental data beyond SPR was provided to better inform participants, including solubility and aggregation measured by DLS, binding by SPR to an unrelated protein (NSD2-PWWP1), and data from orthogonal binding assays (^{19}F-NMR shown here). Data for compound CACHE_1195_6 is shown as an example. (c) Crystal structure of CACHE_1193_26 bound at the interface of two LRRK2 monomers (PDB 9C61), distant from the targeted central cavity, at a site lined by G2385, recurrently mutated to Arg in PD patients.

total, 73 compounds selected by 18 participants had a measurable dissociation constant (K_D) value better than 150 μM and greater than 30% binding (R/R_{max}) (Table S4). To assess whether the binding signal was on target, SPR was used to evaluate binding to an unrelated target, the first PWWP domain of the protein methyltransferase NSD2 (NSD2-PWWP1). Seventeen compounds bound NSD2-PWWP1 with K_D values ranging from 2 to 177 μM.

Some of the 73 compounds showed signs of aggregation or poor solubility as measured by dynamic light scattering (DLS).[16] None of the tested compounds showed clear signs of binding in differential scanning fluorimetry (DSF) or isothermal titration calorimetry (ITC) assays and only two compounds (CACHE_1195_6, SPR K_D 117 μM and CACHE_1210_69, K_D 117 μM) out of 11 tested using a ^{19}F-NMR assay bound to LRRK2-WDR (Figure 4b).

The only successful cocrystallization or crystal soaking attempt was with compound CACHE_1193_26 (SPR K_D 46 μM). The binding pose captured experimentally was not at the central cavity where the compound was docked, but at the interface of two LRRK2-WDR monomers (Figure 4c, Table S5). To test the hypothesis that the binding mode captured in the crystal structure might be induced by crystallization, we generated and purified a mutant form of the LRRK2 WDR domain in which the glycine residue lining the observed pocket was replaced with arginine (LRRK2-WDR G2385R), a space filling residue that is also found in PD patients.[11] Using SPR, we observed that the compound bound equally well to the two forms, strongly suggesting that the binding pocket observed in the crystal structure is distinct from the one exploited in solution.

Some of the 73 SPR hits showed suboptimal behavior in solution, including 37 compounds with signs of poor solubility by DLS, some of which also produced a binding signal against the antitarget NSD2-PWWP1, and almost none were confirmed with an orthogonal biophysical assay. In spite of these red flags, we decided to advance all 73 compounds of interest to the hit expansion stage to avoid false negatives, which we discuss later.

Selection and Experimental Testing of Round 2 Compounds. While the focus of Round 1 was to avoid false negatives, Round 2 was focused on avoiding false positives. Here, participants who had predicted one of the 73 compounds advanced from Round 1 selected up to 50 commercially available analogs of those compounds for Round 2. The aim of Round 2 was to generate structure activity relationship (SAR) to build confidence that binding signals were not artifacts from the assay or driven by other irrelevant factors (e.g., aggregation).

A total of 714 compounds were selected by participants for experimental testing in Round 2, representing 23 to 49 compounds per participant and up to 43 analogs per parent molecule (Table 1, Table S6). Because participants with more than one confirmed hit in Round 1 tended to submit analogs of their strongest hit, and due to the lack of commercial availability of some analogs, only analogs of 42 of the total 73 Round 1 hits were tested in Round 2. As in Round 1, SPR was the primary assay. Sixty-one compounds had a measurable K_D value (8.5% hit-rate) with acceptable SPR parameters (maximum binding signal (R_{max}) > 30% of the expected signal, $T(K_D) > 1$ and $Chi^2 <$ 10% R_{max}), 31 of which had a K_D < 150 μM (Table S3). ^{19}F-NMR and DSF assays were used to orthogonally confirm SPR hits.

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Summary of Round 2 Experimental Results

Participant	Round 1 compound of interest	Analogs tested in Round 2	Analogs with K_D < 150 μM	Best K_D [μM]	Confirmed by ^{19}F-NMR or DSF
1179	CACHE_1179_36	34	2	47	
1179	CACHE_1179_94	7	1	13	
1181	CACHE_1181_33	32	2	56	Yes
1181	CACHE_1181_50	1	1	147	
1183	CACHE_1183_2	4	0	0	
1183	CACHE_1183_25	43	3	5	Yes
1184	CACHE_1184_17	23	0	0	
1184	CACHE_1184_22	21	0	0	
1186	CACHE_1186_2	36	0	0	
1186	CACHE_1186_70	2	0	0	
1186	CACHE_1186_85	9	0	0	
1187	CACHE_1187_59	23	1	93	
1188	CACHE_1188_3	13	0	0	
1188	CACHE_1188_48	15	0	0	
1188	CACHE_1188_92	21	1	150	
1193	CACHE_1193_26	25	2	66	Yes
1193	CACHE_1193_27	13	1	140	
1193	CACHE_1193_8	1	0	0	
1195	CACHE_1195_1	3	0	0	
1195	CACHE_1195_28	17	0	0	
1195	CACHE_1195_32	3	0	0	
1195	CACHE_1195_43	8	0	0	
1195	CACHE_1195_5	5	0	0	
1195	CACHE_1195_6	5	1	58	Yes
1195	CACHE_1195_60	3	1	64	
1200	CACHE_1200_39	11	0	0	
1200	CACHE_1200_51	25	0	0	
1201	CACHE_1201_96	38	2	59	
1202	CACHE_1202_13	13	2	93	Yes
1202	CACHE_1202_37	5	1	37	
1202	CACHE_1202_45	6	1	88	
1202	CACHE_1202_80	1	0	0	
1202	CACHE_1202_97	7	0	0	
1203	CACHE_1203_35	25	0	0	
1205	CACHE_1205_40	34	0	0	
1205	CACHE_1205_93	12	0	0	
1207	CACHE_1207_98	49	0	0	
1208	CACHE_1208_88	48	1	82	
1208	CACHE_1208_98	1	1	133	
1209	CACHE_1209_15	37	5	65	Yes
1209	CACHE_1209_48	4	0	0	
1210	CACHE_1210_69	35	2	71	Yes

All data were evaluated by an independent Hit Evaluation Committee composed of industry experts (Table S1). Overall, seven chemical series were convincingly confirmed with two orthogonal assays (Tables 1, S3, S7, Figure 5). These are the first reported molecules targeting LRRK2-WDR. Other chemical series with a lower score from the Hit Evaluation Committee may also be valid LRRK2-WDR binders. Interestingly, some compounds of interest that displayed significant liability in Round 1 produced convincing chemical series in Round 2 (Figure 5). For instance, CACHE_1181_33 (K_D 123 μM) showed signs of insolubility and aggregation at 200 μM as measured by DLS, but its fluorinated analog, CACHE-HO_1181_24 (K_D 56 μM), was soluble, did not aggregate at 200 μM and showed a clear binding signal by ^{19}F-NMR. This result supports the decision to take an inclusive approach to advancing nonconvincing compounds of interest from Round 1 to Round 2 in order to avoid false negatives.

Additionally, we cannot discount the possibility that some of the chemical series that were not validated by ^{19}F-NMR (due to lack of a fluorinated compound) or by DSF (possibly due to distinct mode of binding) in Round 2 may still be valid LRRK2-WDR ligands. As such, CACHE results should be interpreted as evidence that certain computational workflows are performing well, but do not necessarily imply that other workflows are performing poorly. With this in mind, our next step was to analyze common and distinct features and design strategies adopted by the best performing CACHE participants.

Emerging Trends from the Seven Best Performing Computational Workflows. Superimposing the docked poses of some of the top hits reveals that, while all were predicted to occupy the central channel of the LRRK2-WDR domain, there is no significant overlap in the predicted network of interactions with the protein, reflecting the open-ended and challenging

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

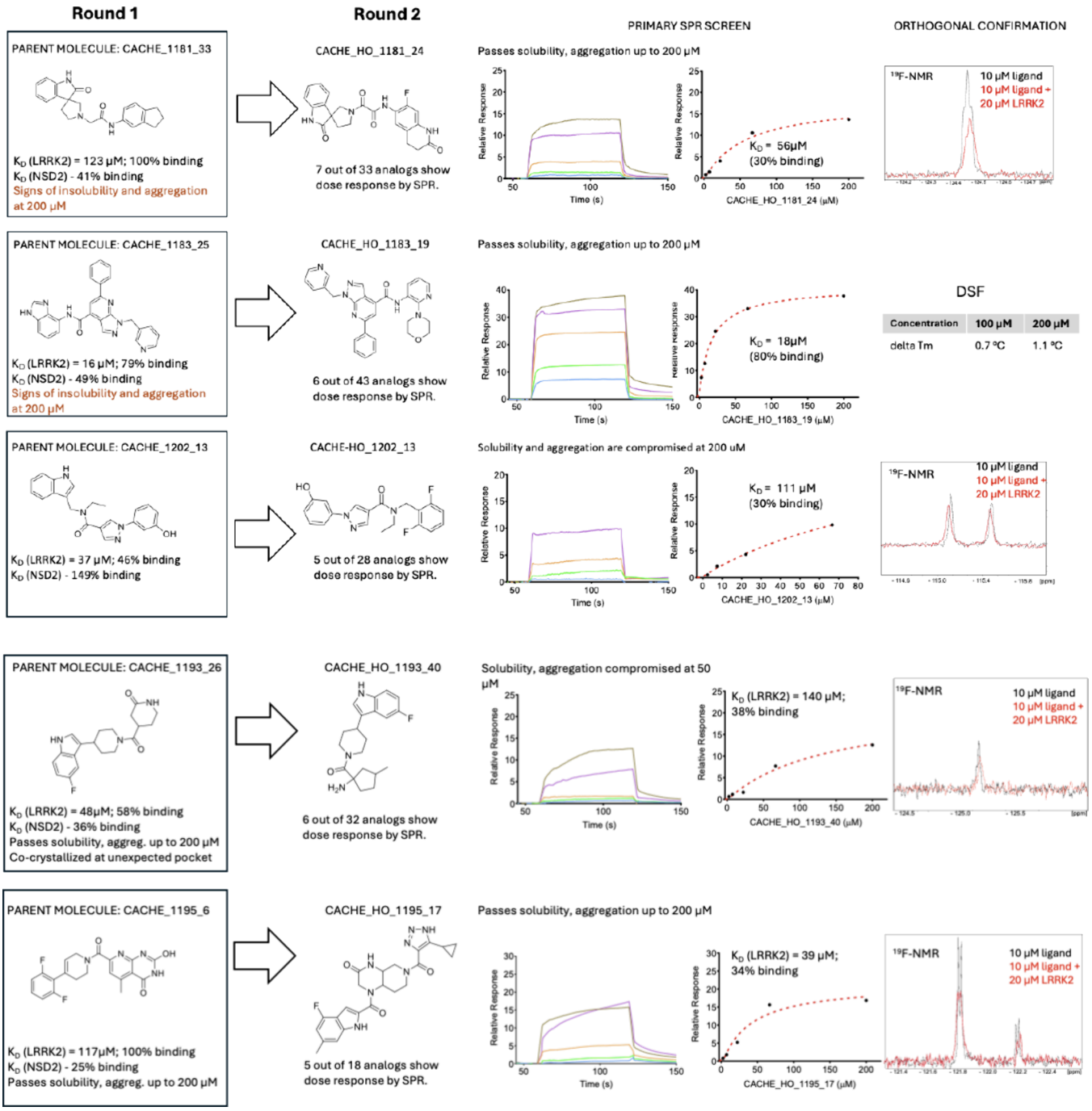


Figure 5. continued

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

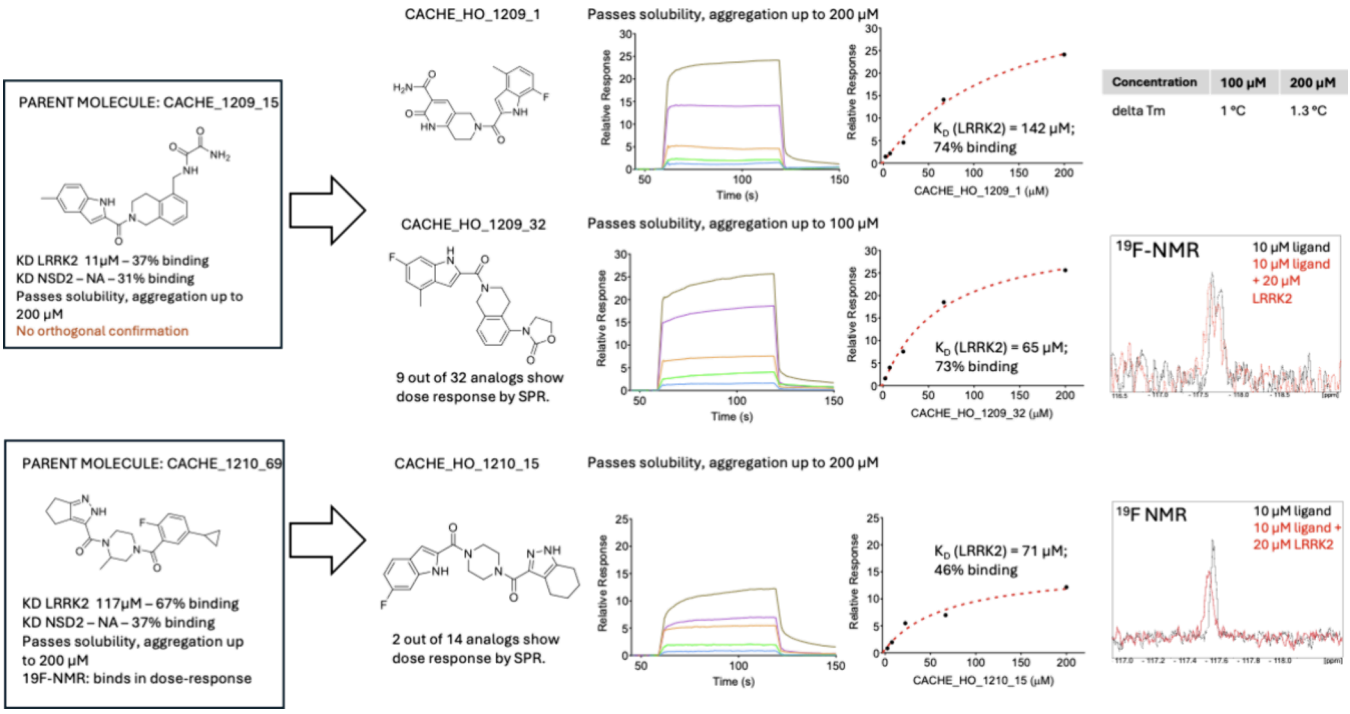


Figure 5. Top seven chemical series identified in Round 2. Activity of the parent molecules and experimental data from Round 2 analogs are shown, including SPR sensorgrams, ^{19}F-NMR spectra and thermal shifts from DSF. Computational workflow IDs are encoded into compound names.

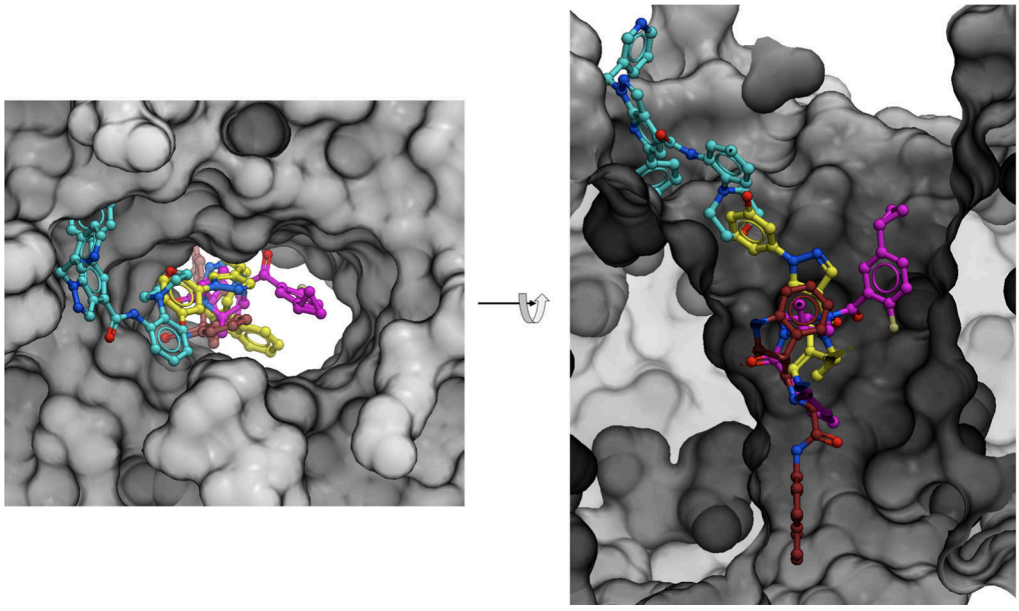


Figure 6. Docked poses of experimental hits. Four compounds are shown: CACHE_1183_13 (light blue), CACHE_1202_13 (yellow), CACHE_1210_69 (pink), CACHE_1181_33 (maroon). Top scoring poses for each ligand are shown. Computational workflows are included in the compound names and summarized in Figure 2.

nature of this binding site for structure-based drug design (Figure 6).

The seven computational workflows that produced a chemical series experimentally confirmed in Round 2 were highly diverse (Figure 7, detailed description in https://cache-challenge.org/challenges/predict-hits-for-the-wdr-domain-of-lrrk2/computational-methods), though close examination makes a few recurring trends and strategies apparent.

First, all but one of these workflows included at least one ML step, and five used some element of deep learning. Workflow 1181 (WF1181) adopted a physics-based high-throughput docking strategy complemented with a 3D convolutional neural network (CNN) scoring function implemented in GNINA[17] to select compounds. WF1209 used DeepDocking,[18,19,26] where a deep neural network (DNN) predicts docking scores in order to rapidly screen an ultralarge library, followed by more refined active learning selection cycles where free energy calculation data was used to train an ML model.[19] WF1193 used Glide docking scores (Schrödinger, New York) to train REINVENT, a recurrent neural network (RNN) and transformer-based ML

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

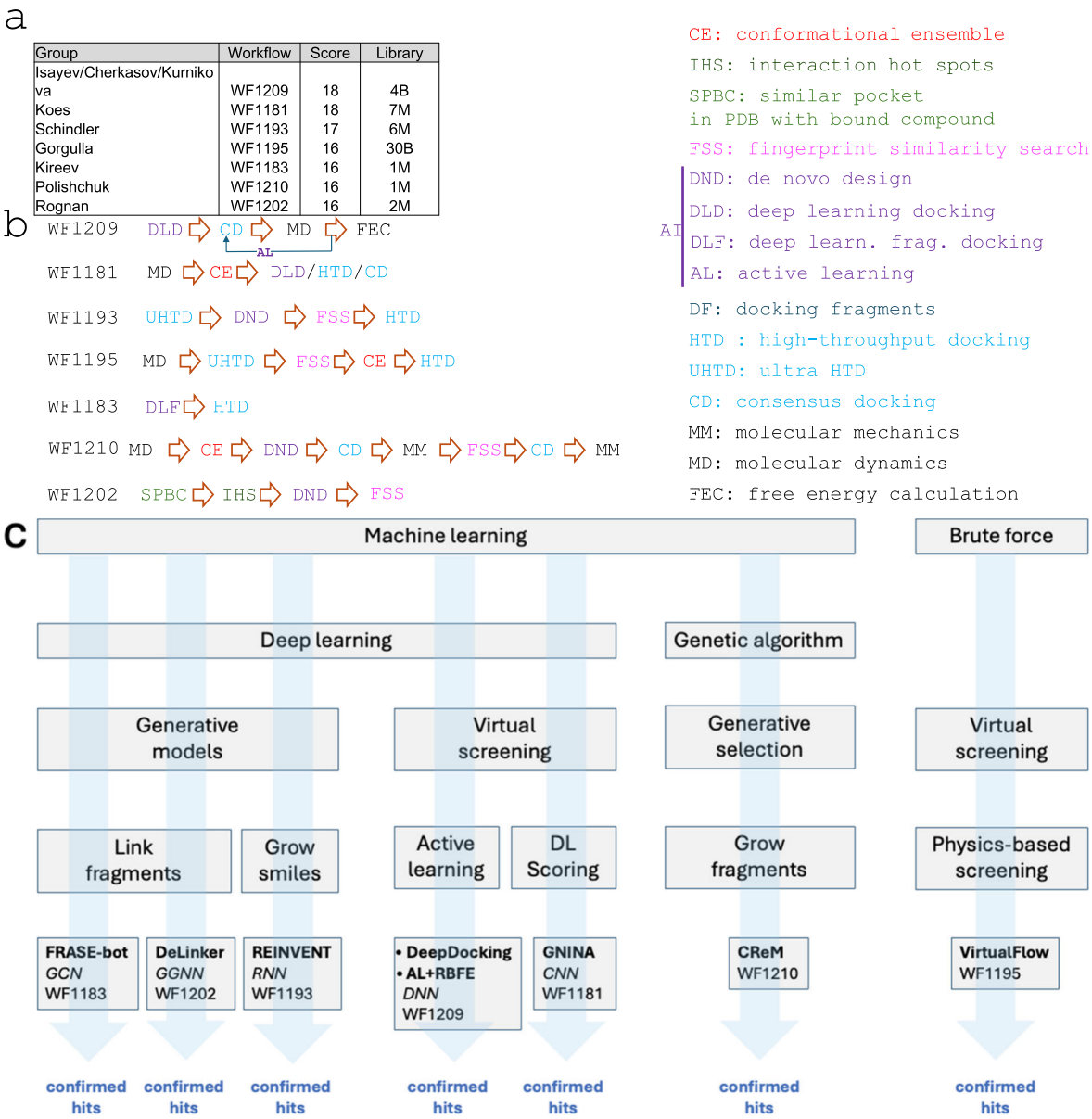


Figure 7. Overview of the best performing computational workflows. Details of the seven computational workflows that had a chemical series experimentally confirmed with two binding assays. (a) Team leads, workflow IDs, aggregated scores from the hit evaluation committee, and size of the library originally screened (Table S1 − Each committee member gave a score from 0 to 5 to each workflow based on the experimental data). (b) Schematics representation of the computational steps for each workflow. (c) Classification of the workflows based on some of their distinct features. Defining computational tools are outlined in bold. Neural network architectures are shown in italic. CNN: convolutional neural network; DNN: deep neural network; GCN: graph convolutional network; GGNN: gated graph neural network; RNN: recurrent neural network. DL: deep learning.

model[20,21] that generated de novo ligand candidates. WF1183 and WF1202 decomposed the binding site into local protein microenvironments with ligand-occupied structural homologues in the PDB. WF1183 used FRASE-bot[22] where a graph convolutional network (GCN) distills an optimal feature vector from the protein−ligand interaction graphs to select the best fragments. Commercial molecules overlapping fragment pairs were then docked and ranked. WF1202 applied the POEM screening cascade where fragments positioned with a point cloud pocket registration system are linked with DeLinker, a generative ML model with a multimodal encoder-decoder setup based on a standard gated graph neural network (GGNN).[23−25] These successful strategies used deep learning to accelerate classical methods (WF1209),[19,26] to predict affinity (WF1181), or to generate new molecules (WF1183, WF1193, WF1202). Finally, WF1210 did not use a neural network architecture but used CReM[27] to grow previously docked fragments, followed by fine-tuning using a genetic algorithm. Workflow 1195 (WF1195) used VirtualFlow[28] (first generation) to deploy conventional physics-based virtual screening tools across tens of thousands of CPUs to efficiently dock an ultralarge chemical library.

Three of the workflows (WF1183, WF1202 and WF1210) used fragment-based approaches. Three used some form of de novo generative method to invent molecules customized for the target site (WF1193, WF1202, WF1210) followed by fingerprint similarity search to identify commercially available chemical analogs. As no ligand was reported at the outset of this challenge, available structures of the binding pocket were in the apo state, which is typically challenging for ligand binding and virtual

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

screening. Three design strategies included molecular dynamics simulations to generate a conformational ensemble of the target site against which compounds were docked (WF1181, WF1195, WF1210).

Together, these results demonstrate that multiple design strategies and technical tools can successfully drive the structure-based discovery of pioneer ligands for an unprecedented target. Significant differences were also observed in the amount of computational resources used (Table S8). Two of the seven best performing workflows exclusively deployed conventional computational tools and methodologies that have been in use for decades (WF1195, WF1210), achieving results comparable to those obtained with deep learning-driven screening cascades. This indicates that advanced neural network architectures did not lead to a breakthrough in this challenge.

■ DISCUSSION

Unlike previous computational challenges, where participants were asked to predict pregenerated experimental data blinded to them, CACHE is the first benchmarking challenge where computational predictions are experimentally tested prospectively. A new CACHE challenge is launched every four months, each addressing a different type of technical challenge (e.g., availability of protein structures and known ligands). For each CACHE target, suitable assays are used to confirm predicted hits.

In this first iteration, the selected target had no known ligand to validate computational workflows or experimental assays. While this high bar may be considered a debatable choice for an inaugural target, we believe it sets reasonable expectations for nonexperts regarding virtual screening. Despite the difficulty presented by the chosen target, seven independent teams were able to use the apo structure of LRRK2-WDR to predict ligands, which were subsequently confirmed as mid-to-high micromolar binders through experimental testing.

Several important lessons emerged from CACHE #1, spanning both experimental and in-silico aspects of the challenge. One major challenge faced by the experimental team was the poor solubility of many of the predicted molecules. Virtual screening can yield high hit rates and potent molecules when experienced researchers work with well-characterized targets, where the structural chemistry of the target or target-class is well-understood, and known ligands are available to identify favorable pocket conformation(s), define interaction hotspots, and validate docking protocols.[29−31] However, hit rates are typically low or null and compounds weak when computational screening is applied to underexplored proteins with no known ligand, as was the case here. As a result, predicted molecules must be tested at high concentrations (up to 200 μM in this challenge), where they frequently precipitate or aggregate. Indeed, 53% of the molecules tested by DLS in Round 1 were not fully soluble at 200 μM in the SPR buffer minus detergent.

To our knowledge, solubility prediction, when not trained on a given chemical series, remains unreliable. However, introducing a mechanism to filter out poorly soluble compounds before procurement and testing would improve the screening process. This would also reduce uncertainties related to compounds of interest showing weak activity and poor solubility.

In Round 1, we chose not to filter out compounds that behaved poorly in solution, for instance due to low solubility. Unlike a typical drug discovery project, the CACHE experimental team cannot afford to disregard second tier hit candidates. It is important not to prematurely dismiss a computational pipeline that may have generated structurally valid molecules. As a result, some dubious molecules were advanced to Round 2, where the focus is to identify convincing hits and clearly successful computational pipelines. In some cases, these successful pipelines emerged despite their producing problematic compounds in Round 1 (e.g.: Figure 5, WF1181, WF1183).

While this challenge succeeded in providing a unified metric for comparing computational screening pipelines and highlighting successful ones, there are areas for future improvement. First, although we identified methods that performed well, we cannot definitively conclude that others did not. Some workflows that did not rank in the top seven produced chemically related hits detected by SPR but not confirmed by DSF or by ^{19}F-NMR. It is possible that some of these hits were valid but did not produce a detectable binding signal in the orthogonal assays.

Second, the original white paper detailing the scope and operational setup of the CACHE challenges[1] proposed a step where all participants would blindly screen a library composed of all compounds predicted in Round 1. This would have allowed direct comparison of methods using the same compound library. However, with few experimentally confirmed hits and only two or fewer hits predicted by each participant, this data was insufficient for a statistically significant analysis, and the step was ultimately dismissed.

Third, structure-based virtual screening is not an exact science and the same computational workflow may succeed for one researcher and fail for another. While this variability may not be an issue when the goal is to identify suitable partners for drug discovery projects (i.e., a successful combination of team and technology), it could be seen as a limitation when evaluating scientific methodologies. One perspective, related to the first point, is that CACHE remains a valuable metric for identifying successful workflows: experimental validation of active molecules implies that the workflow produced valid hits, and humans selected some of them. Conversely, if a pipeline fails to produce valid molecules, even the most skilled computational or medicinal chemists would struggle to identify active compounds based on intuition alone. To eliminate the human factor, participants could submit containerized versions of their pipelines instead of a set of predicted molecules. CACHE organizers would then run these methods blindly and select compounds. A similar approach has been used in the CELPP challenges[5] and could enhance objectivity here. Alternatively, CACHE participants may be asked in the future to clearly specify the human factor in compound selection.

■ CONCLUSION

The first iteration of the CACHE challenges closely followed a process carefully designed by a diverse group of stakeholders in computational hit finding.[1] Despite expected and unexpected challenges encountered along the way, CACHE #1 successfully identified computational pipelines in structure-based virtual screening and produced the first experimentally confirmed ligands binding outside of the kinase domain of LRRK2, an important target in Parkinson's disease. These ligands could serve as a starting point to explore previously untested therapeutic hypotheses. The range of methods used, including many that leverage modern neural network architectures, reflects an intensely dynamic and explorative community. However, despite the current hype surrounding AI-driven

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

drug discovery, a true breakthrough in the field has yet to emerge.

METHODS

Computational Workflows. Computational methods are available from https://cache-challenge.org/results-cache-challenge-1

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.[32] 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 min at 10 °C), 72−96 h postinfection with well-developed signs of infections and 70−80% viability as previously described.[30] Harvested cells were resuspended in 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5 mM imidazole and 5% glycerol, 1X protease inhibitor cocktail (100 X protease inhibitor stock in 70% ethanol (0.25 mg/mL Aprotinin, 0.25 mg/mL Leupeptin, 0.25 mg/mL Pepstatin A and 0.25 mg/mL E-64) or Pierce Protease Inhibitor Mini Tablets, EDTA-free. The cells were lysed chemically by addition of 1 mM PMSF, 1 mM 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 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% glycerol, containing 5 mM, 15 mM and 250 mM imidazole, respectively. The eluted proteins were then supplemented with 2 mM 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 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 2 mM 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 20 mM Tris-HCl, pH 7.4, containing 300 mM NaCl, 2 mM 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. The binding affinity of compounds was assessed by Surface plasmon resonance (SPR, Biacore 8K, Cytiva Inc.) at 25 °C. Biotinylated LRRK2 (2141−2527aa - https://www.addgene.org/210899/) 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 compound was 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 multicycle kinetics with a 60 s contact time and a dissociation time of 120 s at a flow rate of 40 μL/min. Kinetic 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. LRRK2 was diluted to 0.1 mg/mL in buffer (100 mM Hepes, 100 mM NaCl, pH 7.5) in the presence of 5x SYPRO Orange dye (Life Technologies, S-6650) and serially titrated compounds (up to 200 μM) in a total volume of 20 μL in a white polypropylene 384-well plate (Axygen, PCR-384-LC480-W). DSF was performed in a LightCycler 480 II (Roche Applied Science, Penzberg, Germany) using a 4 °C/min temperature gradient from 20 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.[33]

Dynamic Light Scattering. The solubility of compounds was estimated by DLS that directly measures compound aggregates and laser power in solution. Compounds were serially diluted directly from DMSO stocks, then diluted 50x into filtered 10 mM HEPES pH 7.4, 150 mM NaCl(2% DMSO final). The resulting samples were then distributed into 384-well plates (black with a clear bottom, Corning 3540), with 20 μL in each well. The sample plate was centrifuged at 3500 rpm for 5 min before loading into DynaPro DLS Plate Reader III (Wyatt Technology) and analyzed as previously described.[34,35]

^{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 (pH 7.4, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, and with 5% D_2O). 1D-^{19}F spectra were collected at 298 K on a Bruker AvanceIII 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 Topspin 3.5.

Crystallization and Structural Determination. Human LRRK2 WDR domain (residues 2142−2527) was expressed, purified and crystallized as described previously (PMID: 30635421). Apo-LRRK2 WDR domain crystals were obtained by mixing equimolar amounts of protein (concentrated at 9 mg/mL) and precipitant solution containing 0.1 M Tris-HCl at pH 8.5, 1 M LiCl, 14% (w/v) polyethylene glycol (PEG) 6000, and 10% galactose in a manual plate vapor-diffusion hanging drops. LRRK2 crystals were then soaked into a 1 μL reservoir solution supplemented with 1 mM CACHE 1193−26 (dissolved from a previously prepared 100 mM DMSO stock solution) and 10% (v/v) Ethylene glycol for 2 h at room temperature, then mounted and cryo-cooled in liquid nitrogen. Diffraction data were collected at the 24ID-E beamline at the Advanced Photon Source (APS). Data set was processed with HKL3000.[36] Initial phases were obtained by using Apo-LRRK2 WDR domain (PDB ID:6DLO) as initial model in Fourier transform with refmac5.[37] Model building was performed in COOT[38] and the structure

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

was validated with Molprobity.[39] CACHE 1193−26 structure restraints were generated using grade Web Server (http://grade.globalphasing.org).

■ ASSOCIATED CONTENT

Data Availability Statement

The crystal structure of LRRK2-WDR in complex with CACHE_1193_26 was deposited in the Protein Data Bank, PDB code 9C61. All files and document available from the CACHE#1 data release webpage and the description of computational methods posted on the CACHE Web site are also posted on Zenodo at https://zenodo.org/records/13820554 (DOI 10.5281/zenodo.13800102), including manuals on interpreting SPR and DSF data for nonexperts.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c01267.

CACHE1 committees (Table S1), Round 1 compounds (Table S2), Experimental data for all compounds (Table S3), Round 1 compounds advanced to Round 2 (Table S4), X-ray data collection (Table S5), Round 2 compounds (Table S6), Score from Hit Evaluation Committee (Table S7), and Computational resources (Table S8) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Matthieu Schapira − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada;* orcid.org/0000-0002-1047-3309; Email: matthieu.schapira@utoronto.ca

Authors

Fengling Li − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*

Suzanne Ackloo − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada;* orcid.org/0000-0002-9696-1839

Cheryl H. Arrowsmith − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2C4, Canada*

Fuqiang Ban − *Vancouver Prostate Centre, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada*

Christopher J. Barden − *Treventis Corporation, Toronto, Ontario M5T 0S8, Canada; University Health Network, Toronto, Ontario M5G 2C4, Canada*

Hartmut Beck − *Bayer AG, Drug Discovery Sciences, 42096 Wuppertal, Germany;* orcid.org/0000-0002-7826-8467

Jan Beránek − *Department of Biochemistry and Microbiology, University of Chemistry and Technology, Technická 5 16628 Prague, Czech Republic;* orcid.org/0000-0002-5347-3966

Francois Berenger − *Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan;* orcid.org/0000-0003-1377-944X

Albina Bolotokova − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*

Guillaume Bret − *Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France;* orcid.org/0000-0001-5434-0906

Marko Breznik − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*

Emanuele Carosati − *Department of Chemical and Pharmaceutical Sciences, University of Trieste, 34127 Trieste, Italy;* orcid.org/0000-0003-0657-5035

Irene Chau − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*

Yu Chen − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*

Artem Cherkasov − *Vancouver Prostate Centre, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada;* orcid.org/0000-0002-1599-1439

Dennis Della Corte − *Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, United States;* orcid.org/0000-0002-8884-9724

Katrin Denzinger − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany;* orcid.org/0000-0003-4222-8161

Aiping Dong − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*

Sorin Draga − *Virtual Discovery, Inc., Boston, Massachusetts 02108, United States; Non-Governmental Research Organization Biologic, Bucharest 032044, Romania*

Ian Dunn − *Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States*

Kristina Edfeldt − *Structural Genomics Consortium, Department of Medicine, Karolinska University Hospital and Karolinska Institutet, 171 76 Stockholm, Sweden*

Aled Edwards − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Conscience Medicines Network, Toronto, Ontario M5G 1L7, Canada;* orcid.org/0000-0002-4782-6016

Merveille Eguida − *Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France;* orcid.org/0000-0002-0976-0239

Paul Eisenhuth − *Institute for Drug Discovery, Medical Faculty, Leipzig University, Leipzig, Saxony 04103, Germany; Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Leipzig, Saxony 04105, Germany*

Lukas Friedrich − *Computational Drug Design, Merck KGaA, 64293 Darmstadt, Germany*

Alexander Fuerll − *Institute for Drug Discovery, Medical Faculty, Leipzig University, Leipzig, Saxony 04103, Germany;* orcid.org/0009-0008-3141-2434

Spencer S Gardiner − *Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, United States*

Francesco Gentile − *Vancouver Prostate Centre, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada; Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada; Ottawa Institute of Systems Biology, University of Ottawa, K1H 8M5 Ottawa, Ontario, Canada*

Pegah Ghiabi − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*

Elisa Gibson − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada;* orcid.org/0000-0002-7112-337X

Marta Glavatskikh − *University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States*

Christoph Gorgulla − *St. Jude Children's Research Hospital, Memphis, Tennessee 38105, United States; Department of Physics, Harvard University, Cambridge, Massachusetts 02138, United States*
Judith Guenther − *Bayer AG, Drug Discovery Sciences, 13353 Berlin, Germany;* orcid.org/0000-0001-5794-8984
Anders Gunnarsson − *Structure and Biophysics, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Mölndal 431 50, Sweden*
Filipp Gusev − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-1167-345X
Evgeny Gutkin − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0003-4522-6049
Levon Halabelian − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada;* orcid.org/0000-0003-4361-3619
Rachel J. Harding − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada; Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada; Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario M5S 3M2, Canada;* orcid.org/0000-0002-1134-391X
Alexander Hillisch − *UCB BioSciences GmbH, 40789 Monheim am Rhein, Germany*
Laurent Hoffer − *Drug Discovery, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada*
Anders Hogner − *Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D,, AstraZeneca, 431 50 Gothenburg, Sweden;* orcid.org/0000-0003-3823-4534
Scott Houliston − *Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2C4, Canada*
John J Irwin − *Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States*
Olexandr Isayev − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0001-7581-8497
Aleksandra Ivanova − *Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University in Olomouc, 77900 Olomouc, Czech Republic*
Celien Jacquemard − *Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France*
Austin J Jarrett − *Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, United States;* orcid.org/0000-0001-8522-0309
Jan H. Jensen − *Department of Chemistry, University of Copenhagen, 2100 Copenhagen, Denmark;* orcid.org/0000-0002-1465-1010
Dmitri Kireev − *Department of Chemistry, University of Missouri, Columbia, Missouri 65211-7600, United States;* orcid.org/0000-0001-8479-8555
Julian Kleber − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany;* orcid.org/0000-0001-5518-0932
S. Benjamin Koby − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*
David Koes − *Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-6892-6614
Ashutosh Kumar − *Center for Biosystems Dynamics Research, RIKEN, Yokohama, Kanagawa 230-0045, Japan;* orcid.org/0000-0003-3754-8996
Maria G. Kurnikova − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-8010-8374
Alina Kutlushina − *Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University in Olomouc, 77900 Olomouc, Czech Republic;* orcid.org/0000-0003-3496-7984
Uta Lessel − *Boehringer Ingelheim Pharma GmbH & Co. KG, 88400 Biberach an der Riss, Germany;* orcid.org/0000-0002-1698-6017
Fabian Liessmann − *Institute for Drug Discovery, Medical Faculty, Leipzig University, Leipzig, Saxony 04103, Germany*
Sijie Liu − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Wei Lu − *Galixir Technologies, 200100 Shanghai, China*
Jens Meiler − *Institute for Drug Discovery, Medical Faculty, Leipzig University, Leipzig, Saxony 04103, Germany; Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Leipzig, Saxony 04105, Germany; Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States*
Akhila Mettu − *Department of Chemistry, University of Missouri, Columbia, Missouri 65211-7600, United States*
Guzel Minibaeva − *Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University in Olomouc, 77900 Olomouc, Czech Republic;* orcid.org/0000-0001-7964-4842
Rocco Moretti − *Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States*
Connor J Morris − *Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, United States*
Chamali Narangoda − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*
Theresa Noonan − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Leon Obendorf − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Szymon Pach − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Amit Pandit − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Sumera Perveen − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*
Gennady Poda − *Drug Discovery, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada; Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario M5S 3M2, Canada*

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Pavel Polishchuk − *Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University in Olomouc, 77900 Olomouc, Czech Republic*
Kristina Puls − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Vera Pütter − *Nuvisan ICB GmbH, 89231 Neu-Ulm, Germany*
Didier Rognan − *Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France;* orcid.org/0000-0002-0577-641X
Dylan Roskams-Edris − *Conscience Medicines Network, Toronto, Ontario M5G 1L7, Canada;* orcid.org/0000-0002-4526-2296
Christina Schindler − *Computational Drug Design, Merck KGaA, 64293 Darmstadt, Germany;* orcid.org/0000-0002-8980-048X
François Sindt − *Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France*
Vojtěch Spiwok − *Department of Biochemistry and Microbiology, University of Chemistry and Technology, Technická 5 16628 Prague, Czech Republic*
Casper Steinmann − *Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark;* orcid.org/0000-0002-5638-1346
Rick L. Stevens − *Department of Computer Science, University of Chicago, Chicago, Illinois 60637, United States; Argonne National Laboratory, Lemont, Illinois 60439, United States*
Valerij Talagayev − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany*
Damon Tingey − *Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, United States*
Oanh Vu − *Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States;* orcid.org/0000-0003-4704-7538
W. Patrick Walters − *Relay Therapeutics, Cambridge, Massachusetts 02141, United States;* orcid.org/0000-0003-2860-7958
Xiaowen Wang − *Department of Chemistry, University of Missouri, Columbia, Missouri 65211-7600, United States*
Zhenyu Wang − *Global Institute of Future Technology, Shanghai Jiao Tong University, 200240 Shanghai, China; Galixir Technologies, 200100 Shanghai, China*
Gerhard Wolber − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-5344-0048
Clemens Alexander Wolf − *Computational Molecular Design, Institute of Pharmacy, Freie Universitaet Berlin, 14195 Berlin, Germany;* orcid.org/0000-0002-5682-1815
Lars Wortmann − *Boehringer Ingelheim Pharma GmbH & Co. KG, 88400 Biberach an der Riss, Germany;* orcid.org/0000-0001-6514-947X
Hong Zeng − *Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada*
Carlos A. Zepeda − *Treventis Corporation, Toronto, Ontario M5T 0S8, Canada*
Kam Y. J. Zhang − *Center for Biosystems Dynamics Research, RIKEN, Yokohama, Kanagawa 230-0045, Japan;* orcid.org/0000-0002-9282-8045
Jixian Zhang − *Galixir Technologies, 200100 Shanghai, China*
Shuangjia Zheng − *Global Institute of Future Technology, Shanghai Jiao Tong University, 200240 Shanghai, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c01267

Notes

The authors declare the following competing financial interest(s): C. Gorgulla is a cofounder and consultant of Virtual Discovery, Inc., which is a fee-for-service company for computational drug discovery and a cofounder of Quantum Therapeutics, Inc., a company that uses computational methods for drug development. A. Hillisch is employee of UCB and holds shares in UCB and Bayer AG. L. Wortmann is an employee of Boehringer Ingelheim Pharma GmbH & Co. KG.

■ ACKNOWLEDGMENTS

We thank Claudia Gordijo, Maxwell Morgan, and Richard Gold who played instrumental roles towards the successful launch of the CACHE challenges, and Shabbir Ahmad and Stuart Green for technical discussions on SPR assays. We thank Dr. Hao Wu for sharing the LRRK2 WDR domain construct for structural studies. We thank the staff at the Northeastern Collaborative Access Team, which is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). Experimental testing was supported by an Open Science Drug Discovery grant from Canada's Strategic Innovation Fund (SIF Stream 5) administered by Conscience and the Michael J Fox Foundation, and conducted at the Structural Genomics Consortium, 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], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, and Takeda. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 875510. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Ontario Institute for Cancer Research, Royal Institution for the Advancement of Learning McGill University, Kungliga Tekniska Hoegskolan, Diamond Light Source Limited. This communication reflects the views of the authors and the JU is not liable for any use that may be made of the information contained herein. The Eiger 16M detector on the 24-ID-E beamline is funded by an NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The Koes team's work was supported by grant R35GM140753 from the National Institute of General Medical Sciences. The Kireev team's work was supported by a startup fund provided by the University of Missouri & would like to acknowledge the University of Missouri, Division of IT, Research Computing Support Services for the use of the computing resources. The team led by O. Isayev, A. Cherkasov, and M. Kurnikova acknowledges support by the NSF grants CHE-2154447, DMS-1563291, MCB-1818213, and NIH grant R01NS083660, and would further like to acknowledge that art of the computation for their work was performed at the San Diego Supercomputer Center (SDSC) and the Pittsburgh Supercomputing Center (PSC) through allocation CHE200122 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296 and Frontera computing

project at the Texas Advanced Computing Center (NSF OAC-1818253) award. P. Eisenhuth, member of the joint Vanderbilt & University Leipzig team, would like to acknowledge financial support from the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the program Center of Excellence for AI-research Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig, project identification number: ScaDS.AI. P.E.'s position is funded through an award by ScaDS.AI. This work was further funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through SPP2363 (460865652). J. Meiler acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG) through SFB1423 (421152132), SFB 1052 (209933838), and SPP 2363 (460865652). J.M. is supported by a Humboldt Professorship of the Alexander von Humboldt Foundation. J.M. is supported by BMBF (Federal Ministry of Education and Research) through the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI). This work is partly supported by the Federal Ministry of Education and Research (BMBF) through DAAD project 57616814 (SECAI, School of Embedded Composite AI). Work in the Meiler laboratory is further supported through the NIH (R01 HL122010, R01 DA046138, R01 AG068623, U01 AI150739, R01 CA227833, R01 LM013434, S10 OD016216, S10 OD020154, S10 OD032234). This work was additionally supported by the BMBF-funded German Network for Bioinformatics Infrastructure (de.NBI). Work by the Polishchuk team was supported by the Ministry of Education, Youth and Sports of the Czech Republic through INTER-EXCELLENCE II LUAUS23262, the e-INFRA CZ (ID:90140, ID:90254), ELIXIR-CZ (LM2018131, LM2023055), CZ-OPENSCREEN (LM2018130, LM2023052) grants and by European and Regional Fund project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868).

■ REFERENCES

(1) 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.

(2) Amezcua, M.; Setiadi, J.; Ge, Y.; Mobley, D. L. An Overview of the SAMPL8 Host-Guest Binding Challenge. *J. Comput. Aided Mol. Des* **2022**, *36* (10), 707−734.

(3) Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein-Ligand Pose and Affinity Predictions. *J. Comput. Aided Mol. Des* **2016**, *30* (9), 651−668.

(4) Dunbar, J. B.; Smith, R. D.; Yang, C.-Y.; Ung, P. M.-U.; Lexa, K. W.; Khazanov, N. A.; Stuckey, J. A.; Wang, S.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Selection of the Protein-Ligand Complexes. *J. Chem. Inf Model* **2011**, *51* (9), 2036−2046.

(5) Wagner, J. R.; Churas, C. P.; Liu, S.; Swift, R. V.; Chiu, M.; Shao, C.; Feher, V. A.; Burley, S. K.; Gilson, M. K.; Amaro, R. E. Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking. *Structure* **2019**, *27* (8), 1326−1335.

(6) Steger, M.; Tonelli, F.; Ito, G.; Davies, P.; Trost, M.; Vetter, M.; Wachter, S.; Lorentzen, E.; Duddy, G.; Wilson, S.; Baptista, M. A.; Fiske, B. K.; Fell, M. J.; Morrow, J. A.; Reith, A. D.; Alessi, D. R.; Mann, M. Phosphoproteomics Reveals That Parkinson's Disease Kinase LRRK2 Regulates a Subset of Rab GTPases. *Elife* **2016**, *5*, No. e12813.

(7) Tolosa, E.; Vila, M.; Klein, C.; Rascol, O. LRRK2 in Parkinson Disease: Challenges of Clinical Trials. *Nat. Rev. Neurol* **2020**, *16* (2), 97−107.

(8) West, A. B.; Moore, D. J.; Biskup, S.; Bugayenko, A.; Smith, W. W.; Ross, C. A.; Dawson, V. L.; Dawson, T. M. Parkinson's Disease-Associated Mutations in Leucine-Rich Repeat Kinase 2 Augment Kinase Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (46), 16842−16847.

(9) Berwick, D. C.; Heaton, G. R.; Azeggagh, S.; Harvey, K. LRRK2 Biology from Structure to Dysfunction: Research Progresses, but the Themes Remain the Same. *Mol. Neurodegener* **2019**, *14* (1), 49.

(10) Sanz Murillo, M.; Villagran Suarez, A.; Dederer, V.; Chatterjee, D.; Alegrio Louro, J.; Knapp, S.; Mathea, S.; Leschziner, A. E. Inhibition of Parkinson's Disease-Related LRRK2 by Type I and Type II Kinase Inhibitors: Activity and Structures. *Sci. Adv.* **2023**, *9* (48), No. eadk6191.

(11) 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* (5), 1579−1584.

(12) Schapira, M.; Tyers, M.; Torrent, M.; Arrowsmith, C. H. WD40 Repeat Domain Proteins: A Novel Target Class? *Nat. Rev. Drug Discovery* **2017**, *16* (11), 773−786.

(13) Song, R.; Wang, Z.-D.; Schapira, M. Disease Association and Druggability of WD40 Repeat Proteins. *J. Proteome Res.* **2017**, *16* (10), 3766−3773.

(14) Zheng, S.; Li, Y.; Chen, S.; Xu, J.; Yang, Y. Predicting Drug-Protein Interaction Using Quasi-Visual Questionanswering System. *Nat. Mach Intell* **2020**, *2* (2), 134−140.

(15) Yang, J. J.; Ursu, O.; Lipinski, C. A.; Sklar, L. A.; Oprea, T. I.; Bologa, C. G. Badapple: Promiscuity Patterns from Noisy Evidence. *J. Cheminform* **2016**, *8*, 29.

(16) O'Donnell, H. R.; Tummino, T. A.; Bardine, C.; Craik, C. S.; Shoichet, B. K. Colloidal Aggregators in Biochemical SARS-CoV-2 Repurposing Screens. *J. Med. Chem.* **2021**, *64* (23), 17530−17539.

(17) Francoeur, P. G.; Masuda, T.; Sunseri, J.; Jia, A.; Iovanisci, R. B.; Snyder, I.; Koes, D. R. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *J. Chem. Inf Model* **2020**, *60* (9), 4200−4215.

(18) Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.-T.; Ban, F.; Norinder, U.; Gleave, M. E.; Cherkasov, A. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Cent. Sci.* **2020**, *6* (6), 939−949.

(19) Gusev, F.; Gutkin, E.; Gentile, F.; Ban, F.; Koby, B.; Li, F.; Chau, I.; Ackloo, S.; Arrowsmith, C.; Zeng, H.; Schapira, M.; Isayev, O.; Cherkasov, A.; Kurnikova, M. Active Learning Guided Hit Optimization for the Leucine-Rich Repeat Kinase 2 WDR Domain Based on In Silico Ligand Binding Affinities. *ChemRxiv* **2024**, DOI: 10.26434/chemrxiv-2024-jv0rx.

(20) Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for De Novo Drug Design. *J. Chem. Inf Model* **2020**, *60* (12), 5918−5922.

(21) Loeffler, H. H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L. H.; Engkvist, O. Reinvent 4: Modern AI-Driven Generative Molecule Design. *J. Cheminform* **2024**, *16* (1), 20.

(22) An, Y.; Lim, J.; Glavatskikh, M.; Wang, X.; Norris-Drouin, J.; Hardy, P. B.; Leisner, T. M.; Pearce, K. H.; Kireev, D. In Silico Fragment-Based Discovery of CIB1-Directed Anti-Tumor Agents by FRASE-Bot. *Nat. Commun.* **2024**, *15* (1), 5564.

(23) 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* (20), 13771−13783.

(24) Eguida, M.; Bret, G.; Sindt, F.; Li, F.; Chau, I.; Ackloo, S.; Arrowsmith, C.; Bolotokova, A.; Ghiabi, P.; Gibson, E.; Halabelian, L.; Houliston, S.; Harding, R. J.; Hutchinson, A.; Loppnau, P.; Perveen, S.; Seitova, A.; Zeng, H.; Schapira, M.; Rognan, D. Subpocket Similarity-Based Hit Identification for Challenging Targets: Application to the WDR Domain of LRRK2. *J. Chem. Inf Model* **2024**, *64*, 5344.

(25) 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* (4), 1983−1995.

(26) Gutkin, E.; Gusev, F.; Gentile, F.; Ban, F.; Koby, S. B.; Narangoda, C.; Isayev, O.; Cherkasov, A.; Kurnikova, M. G. In Silico Screening of LRRK2 WDR Domain Inhibitors Using Deep Docking and Free Energy Simulations. *Chem. Sci.* **2024**, *15*, 8800.

(27) Polishchuk, P. CReM: Chemically Reasonable Mutations Framework for Structure Generation. *J. Cheminform* **2020**, *12* (1), 28.

(28) Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature* **2020**, *580* (7805), 663−668.

(29) Kaplan, A. L.; Confair, D. N.; Kim, K.; Barros-Álvarez, X.; Rodriguiz, R. M.; Yang, Y.; Kweon, O. S.; Che, T.; McCorvy, J. D.; Kamber, D. N.; 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-HT2A Receptor Agonists with Antidepressant Activity. *Nature* **2022**, *610* (7932), 582−591.

(30) Alon, A.; Lyu, J.; Braz, J. M.; Tummino, T. A.; Craik, V.; O'Meara, M. J.; Webb, C. M.; Radchenko, D. S.; Moroz, Y. S.; Huang, X.-P.; Liu, Y.; Roth, B. L.; Irwin, J. J.; Basbaum, A. I.; Shoichet, B. K.; Kruse, A. C. Structures of the Σ2 Receptor Enable Docking for Bioactive Ligand Discovery. *Nature* **2021**, *600* (7890), 759−764.

(31) Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224−229.

(32) Hutchinson, A.; Seitova, A. Production of Recombinant PRMT Proteins Using the Baculovirus Expression Vector System. *J. Vis Exp* **2021**, No. 173, DOI: 10.3791/62510.

(33) 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. *Nat. Commun.* **2019**, *10* (1), 5759.

(34) Aleandri, S.; Vaccaro, A.; Armenta, R.; Völker, A. C.; Kuentz, M. Dynamic Light Scattering of Biopharmaceutics-Can. Analytical Performance Be Enhanced by Laser Power? *Pharmaceutics* **2018**, *10* (3), 94.

(35) Allen, S. J.; Dower, C. M.; Liu, A. X.; Lumb, K. J. Detection of Small-Molecule Aggregation with High-Throughput Microplate Biophysical Methods. *Curr. Protoc Chem. Biol.* **2020**, *12* (1), No. e78.

(36) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. HKL-3000: The Integration of Data Reduction and Structure Solution-from Diffraction Images to an Initial Model in Minutes. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 859−866.

(37) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204−1214.

(38) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486−501.

(39) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* **2018**, *27* (1), 293−315.

https://doi.org/10.1021/acs.jcim.4c01267
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX