Drugit: Crowd-sourcing molecular design of non-peptidic VHL binders

AUTHORS

Thomas Scott[1]‡, Christian Alan Paul Smethurst[2]‡, Yvonne Westermaier[2], Moriz Mayer[2], Peter Greb[2], Roland Kousek[2], Tobias Biberger[2], Gerd Bader[2], Zuzana Jandova[2], Philipp S. Schmalhorst[2], Julian E. Fuchs[2], Aniket Magarkar[3], Christoph Hoenke[3], Thomas Gerstberger[2], Steven A. Combs[1], Richard Pape[1], Saksham Phul[1], Sandeepkumar Kothiwale[1], Andreas Bergner[2], Alex G. Waterson[5], Foldit Players†, Harald Weinstabl[2], Darryl B. McConnell[2], Jark Böttcher[2]*, Jens Meiler[1,4]*, Rocco Moretti[1]*

[1] Department of Chemistry, Vanderbilt University, Nashville, TN 37235, USA; Center for Structural Biology, Vanderbilt University, Nashville, TN 37240, USA;

[2] Boehringer Ingelheim RCV, Dr. Boehringer-Gasse 5-11, 1121 Vienna, Austria

[3] Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach an der Riß, Germany

[4] Institute of Drug Discovery, Faculty of Medicine, University of Leipzig, 04103 Leipzig, Germany

[5] Departments of Pharmacology and Chemistry, Vanderbilt University, Nashville, TN 37235, USA; Vanderbilt Institute of Chemical Biology, Nashville TN 37232, USA

‡ Contributed equally

* Corresponding authors

† Consortium. See Supplementary Information for list.

ABSTRACT

Given the role of human intuition in current drug design efforts, crowd-sourced 'citizen scientist' games have the potential to greatly expand the pool of potential drug designers. Here, we introduce ‘Drugit', the small molecule design mode of the online ‘citizen science’ game Foldit. We demonstrate its utility for design with a use case to identify novel binders to the von Hippel Lindau E3 ligase. Several thousand molecule suggestions were obtained from players in a series of 10 puzzle rounds. The proposed molecules were then evaluated by *in silico* methods and by an expert panel and selected candidates were synthesized and tested. One of these molecules, designed by a player, showed dose-dependent shift perturbations in protein-observed NMR experiments. The co-crystal structure in complex with the E3 ligase revealed that the observed binding mode matched in major parts the player’s original idea. The completion of one full design cycle is a proof of concept for the Drugit approach and highlights the potential of involving citizen scientists in early drug discovery.

INTRODUCTION

Despite continued efforts to use rational approaches to automate and accelerate drug development, the endeavor still requires creativity and human intuition. Currently, the role of human intuition is essentially limited to highly trained people in the pharmaceutical industry and a small set of specialist academic groups. Here, we set out to engage a larger non-expert community in the drug design process. To enable this, we extended the Foldit interface with a small molecule design tool ('Drugit'). Foldit is a molecular biology modeling game which has previously been shown to leverage the creative potential of ‘citizen scientists’ to predict and design protein structures[1-4]. Through the gamified Foldit interface, members of the public can suggest protein structure modifications which can be experimentally verified. This is made possible due to cutting edge protein modeling tools with state-of-the-art scoring and incentivization. Likewise, other such games are Phylo[5], GalaxyZoo[6] and Eterna[7]. In these games, 'citizen-scientists' can productively contribute to open research questions, adding value to open innovation together with open science platforms such as SGC’s open science or Boehringer Ingelheim’s opnMe.com. Despite most Foldit players having limited prior design knowledge, the Foldit interface is tailored to encourage productivity in solving design problems, a situation which can be an even greater multiplier for people with prior domain-specific knowledge.

As a test case for engaging the public into the drug design process, we selected the von Hippel Lindau E3-ligase (VHL). VHL is one of the commonly used E3-ligases for proteolysis targeting chimeras (PROTACs), a new drug modality which holds great promise for future medicines. VHL poses an excellent test case as example compounds that bind tightly to VHL exist as do many PROTACs derived from them[8-10]. PROTACs are a rapidly growing segment of therapeutics[11]. Their design is facilitated by their bi-functional nature, combining a target-specific with an E3-ligase specific moiety. These two moieties are mostly independent, and a single E3-ligase moiety can be used with a range of protein-specific moieties to create PROTACs which degrade different target proteins[9]. Currently, the chemical diversity of known VHL-ligands is limited, and most efforts concentrate on the usage of variations of groundbreaking findings by Crews and Ciulli[12]. However, despite making good progress in de-peptidizing parts of the molecule to reduce polarity, the common hydroxyprolinol core crucial for binding affinity imparts unfavorable physicochemical properties due to its polar nature[13,14]. Current molecules are thus limited in their potential use for clinically relevant oral PROTAC drugs due to their poor pharmacokinetic properties, particularly their high efflux as well as poor permeability and stability properties[15]. Having this in mind, the objectives of the puzzle rounds were not only aimed at creating molecules with high affinity to VHL, but also directed by penalties or bonuses of different strength towards topological polar surface area (TPSA), hydrogen bond donor count, and cLogP. Here we describe the Drugit platform, the process we applied to funnel the diverse ideas from the players, the synthesis of selected molecules, and consecutive profiling.

RESULTS

Drugit interface

Foldit design projects proceed as a series of 'puzzles', which are individual competitions lasting about a week. Each puzzle in a series can have a different starting molecule, protein context, scoring details, additional objectives, and tool configuration. Players use the tools available to modify the molecular system and compete to have the best 'score' in each puzzle[1,3]. In small molecule design puzzles, the player is presented with the 3D structure of a starting small molecule docked into the desired target protein binding site, with the goal to optimize the ligand affinity and placement within the protein (Figure 1a). The user can use a set of tools, the primary one being the ’Small Molecule Design’ panel (Figure 1b). This panel includes functionality for the player to add atoms, change element identity, add or delete bonds, change bond order, and add various pre-defined fragments such as phenyl or carboxylate groups. All changes to the molecule are instantaneously checked for basic chemical

feasibility prior to acceptance, and players are informed of grossly unphysical molecules via pop-up messages. Simple rules for protonation state and hydrogen placement are automatically applied to ensure that the ligand is modeled in the correct charge state for physiological conditions. The edited molecule is then aligned with the existing molecule within the current binding pose and the new chemical identity is inserted into the modeled protein.

Molecule Optimization

As the positioning of the ligand within the binding site is critical for compound evaluation, additional tools are available to optimize ligand placement. Both ligand and protein placement can be optimized in the Drugit framework; though for this work, protein atoms were held fixed and only the ligand was allowed to adopt different conformations.

The primary optimization tool is gradient descent minimization within the Rosetta energy function ('wiggle'). To help correct for unfavorable geometries, which can occur from large chemical identity shifts, small molecule design puzzles use a 'dual space' minimization which alternates between both Cartesian and internal coordinate degrees of freedom[16]. Ideal geometries for ligands are enforced by the 'cart_bonded' geometric deviation penalty[17], with ideal geometries derived from the Merck Molecular Force Field (MMFF) values[18].

An alternative Cartesian minimization in the MMFF energy function ('MMFF wiggle') is also available. A limited, fixed protein context is included in the minimization to avoid protein/ligand clashes. The standard Rosetta energy function used in MMFF wiggle may better normalize the structure than the alternative minimization. However, this better normalization may happen at the expense of a decrease in player score.

Conformational sampling of the ligand can be carried out using the 'shake' tool. This currently has limited usefulness, due to the need to generate conformers for the ligands on the fly. While multi-threading allows some calculation of conformers in the background, due to the speed of the ETKDG conformer generation[19] code currently being used, only a small set of conformers are sampled.

A final optimization approach is the manual 'pull' tool, which allows players to grab portions of the molecule and drag them from their current location to a new location in space.

As the tool is intended to be used also by non-expert users and a significant proportion of trial and error in ligand design is foreseen, an ‘undo’ function is implemented. This function tracks multiple saved states, allowing the player to step backwards in their building timeline to either previously created structures of interest, or structures that scored higher than their current iteration (Figure1c).

In-Client Molecule Evaluation

The primary evaluation of binding is the Rosetta energy of the protein and ligand system. This includes the ‘cart_bonded’ term which penalizes ligand with strained internal geometries. The protein/ligand interaction energy can be upweighted to increase the importance of a good binding energy.

As binding energy is not the only consideration for drug candidates, Drugit puzzles make extensive use of 'objectives', i.e., additional bonuses and penalties which are added to the players scores for system properties such as molecular weight, topological polar surface area (TPSA)[20], clogP[21], number of hydrogen bond donors or acceptors, identified structural alerts, or synthetic accessibility[22]. These metrics are currently mostly calculated through the RDKit library[23]. The importance and thresholds of each of these settings can be adjusted on a per-puzzle basis. In addition, most objectives come with a visualization which can highlight those atoms or regions which are causing a sub-optimal score for a particular objective.

VHL Puzzle series

The Drugit VHL Puzzle series was a set of ten puzzles released over consecutive weeks from Oct 20th, 2021 to Jan 12th, 2022. The puzzles were based on the structure of the Von Hippel-Lindau disease tumor suppressor protein in complex with ligand 10, a previously reported binder (PDB code: 5NVX)[8] (Figure 2a). As all reported potent ligands, ligand 10 possesses a central hydroxyproline core motif, mimicking the natural substrate and amide-linked substituents embedded in the peptide binding groove. Players were instructed to both vary the core and find replacements as well as reduce compound polarity, provided that cellular efflux has been identified as a potential limitation of ligand 10-based PROTAC therapeutics. Particularly, the TPSA[20] was used as a proxy for efflux potential. The ten rounds of the puzzle series maintained the protein structure and the overall goals, but varied in their starting ligands, the objectives present, and the weights of their contributions (Supplementary Table S1). An iterative approach was taken in determining puzzle settings. The preliminary results from each round were examined by medicinal chemists, and objectives were added or reweighted to

correct ‘defects’ in the compounds observed. For example, in the first round, no penalty for lipophilicity of the compounds was included, leading to excessively hydrophobic compounds. The addition of an objective which penalized compounds with high clogP immediately resulted in compounds in a more desirable range in the next puzzle (Supplementary Figure S1). Crystallographic waters and additional bonuses for making hydrogen bonds to those waters were also added to the starting structure when medicinal chemists felt including them would improve the quality of the submitted structures.

Over the course of the puzzle series, 333 Foldit players loaded the puzzles and ca. 160 contributed at least one novel compound. The number of distinct compounds submitted per player roughly follows an exponential decay relationship, with the number of compounds per player dropping by roughly one half for every 17 places in the prolificity ranking (Supplementary Figure S2).

Post-Competition Filtering

The Foldit client captures not only the best scoring compound for each player but also allows players to upload compounds which are interesting for later scientific evaluation. Further, the client also takes regular snapshots of each player’s progress. All structures submitted to the server were assembled for analysis and deduplicated on chemical identity. The best scoring structure for each compound was taken as representative. (For compounds present in multiple rounds, structures from later rounds were selected.)

During VHL post-game triaging, ca. 6,500 compounds (Supplementary data file 1; https://fold.it/forum/blog/vhl-ligand-design-updates) were reduced to 19 for expert panel ranking using a combination of computational chemistry tools and medicinal chemistry judgement (Figure 2b).

Upon completion of all Drugit rounds, suggested molecules which had any atoms more than 10 Å from any atom around the starting ligand were first removed. PipelinePilot was then used for automatically filtering out chemically unreasonable molecules among the suggestions made by Drugit players, via simple element counting or property filters (e.g., TPSA, clogP, efflux and permeability predictions, and SMARTS custom filters [see SI for details]).

The 1,073 remaining molecules (Supplementary data file 2) were triaged in SeeSAR[24] as well as Flare[25], retaining 80+ compounds with a total TPSA below 100 $Å^2$ and 20+ compounds with a TPSA between 100 and 120 $Å^2$, each featuring reasonable conformations and no intra- and intermolecular clashes. 60+ compounds with

moderate torsion issues as well as still acceptable clashes, having TPSAs below 100 Å^2 were also selected. Additionally, 10+ compounds from the original game output were rescued and 90+ compounds were slightly modified based on medicinal chemistry knowledge to fix anticipated synthetic issues and high-energy torsional profiles, and to improve affinity.

For a total of 260+ compounds, MDCK efflux, Caco2 efflux, and apical to basal (AB) Caco2 permeability predictions were carried out using *in-house* implementations of various machine learning models embedded in PipelinePilot (see SI for details). Compounds with acceptable predictions for efflux and permeability were then redocked and their overall torsion quality was assessed (see SI for details). Computational chemists then agreed on a set of molecules predicted to make favourable interactions with the protein binding pocket to be taken further. In parallel, medicinal chemists inspected all docking poses and modified certain ideas from Drugit players slightly in anticipation of better physicochemical properties.

In total, 50+ compounds were then submitted to absolute binding free energy calculations (ABFE), including the known VHL binders 2 and 10 as controls. The binding affinity of all these compounds was then calculated by two different methods: FEP+[26-28] as well as the ABFE approach of Biggin and collaborators[29] (see SI for details).

Based on the following criteria, a shortlist of 19 compounds was subsequently selected for assessment by experts: (i) compounds must be easily synthesizable within resource constraints; (ii) compounds must be likely stable, e.g., with functionality represented in marketed compounds; (iii) only one representative per core (i.e., His115/Ser111 alternative binder), with left-hand side (LHS) and right-hand side (RHS) decorations, was allowed; (iv) some chemotypes were excluded based on free energy calculations. After rank-ordering of the compounds by experts, a final compound list with consensus ranking was established for synthesis.

Synthesis and in vitro analysis of designed compounds

The prioritized molecules were evaluated for synthesizability and, in some cases, adapted based on feasibility. To increase redundancy and likelihood of success, precursors and synthetic intermediates were also submitted for testing (Supplementary Table 2). Synthesis routes and data for all 41 molecules generated in this study are described in the Supplementary Information. All synthesized compounds were submitted to a ^{19}F NMR displacement assay, a highly sensitive technique to confirm that the compounds bind to the protein pocket of

interest. Here, a well characterized ^{19}F-containing reporter molecule was used at 50 µM concentration in presence of 2 µM VCB complex[30]. Addition of the compounds to be tested at 500 µM concentration led to different degrees of displacement of the reporter probe from the binding site, and therefore to reappearance of the ^{19}F NMR signal in the spectrum (Figure 3a). Only compound 1 showed a dose-dependent displacement of the ^{19}F NMR probe and was submitted to protein-observed NMR experiments. Protein labeled selectively with ^{13}C methyl groups in the residues Ile, Val, Leu, and Met was used to obtain a K_D = 258 ± 8 µM, by fitting the dose-dependent shifts (Figure 3bc). The competitive behavior was confirmed in a time-resolved fluorescence resonance energy transfer (TR-FRET) assay using a Cy5-labeled VHL Tracer analog[31], revealing displacement of IC_{50} = 264 ± 30 µM. Based on these results, the diastereomeric mixture of compound 1 was co-crystallized in complex with VCB. A crystal structure of the compound bound to VHL was obtained at the resolution of 1.98 Å. The well-resolved electron density allowed the identification of the eutomer. (Figure 4a, Supplementary Table S3, Supplementary Figure S3b). The newly identified hydroxy-piperidinone motif of **1** occupies the hydroxyproline recognition site of VHL, forming conserved hydrogen bonds to Ser-111 and His-115. While the RHS motif is conserved compared to template ligand 10, ring expansion and amide inversion leads to an altered exit vector towards the LHS. Further optimization of compound **1** to engage pi-pi interactions with Trp-88 and Phe-91 from a pyridine substituent can be envisioned (Figure 4ab). Interestingly, the eutomer observed in the co-crystal structures exhibits the (R,R) configuration, not matching the stereochemistry of the original player designed compound with the (R,S) configuration, but does maintain the key pharmacophores (Figure 4cd).

Analysis of compound source

The 19 molecules selected for evaluation came from 9 different players with a range of Foldit experience. Compound selection was performed completely independently and blind of player identity. However, six of the selected compounds came from an experienced medicinal chemist on the evaluation team (C.A.P.S.), who had played the puzzles in his free time, but none of these compounds showed any binding. The compound that showed binding was based on a round 6 design from player 'Nicm25', who joined Foldit in early 2020, and who has no formal medicinal chemistry experience (personal communication).

There was no consistent pattern as to which puzzle round produced the selected compounds. These compounds were generally not the best compounds of the round, nor were they appreciably related to the best-scoring designs for the respective player for that round. Neither the parent compounds nor the as-synthesized

compounds showed an appreciable difference in Rosetta binding energy or docking quality (Supplementary Figure S4).

DISCUSSION

Here, we demonstrate the feasibility of crowd-sourced small molecule design. Given an appropriate starting structure and objectives, game players – including those with limited medicinal chemistry experience – can design compounds which not only successfully meet compound quality criteria, but also bind well to a protein target of interest. Due to the provided objective criteria as well as the post-game property filtering, compound **1** is characterized by its improved predicted physicochemical properties, specifically with respect to the lower TPSA, higher intrinsic permeability, and lower efflux ratio comparised to ligand 10 (Table 1). The identified novel hydroxypiperidinone core as a bioisosteric replacement for the classical hydroxyproline motif provides an interesting starting point for further structural extension by virtue of its chemical nature, though the undecorated 1,3-dicarbonyl motif in turn also presents a configurationally labile stereocenter in its current state. The biophysical affinity of **1** (K_D = 258 ± 8 μM) toward the protein versus the peptide-like starting molecule is significantly lower, however it should be noted that this affinity is averaged due to the diastereomeric mixture. The chemically resolved eutomer would likely compare more favorably against ligand 10 (IC_{50} = 155 nM). Since the starting molecule is the endpoint of an extensive optimization protocol, the player-proposed molecules might better be viewed as an initial promising result of a first design cycle for further structure-activity relationship optimization.

The current design protocol included a fair amount of manual post-processing and evaluation. Torsional favorability played a large role in the selection process, in part due to a lack of explicit consideration during the design. The Drugit client has since been augmented to include scoring for disfavored internal rotations. Less amenable to client incorporation is the hands-on refinement of the player designs by experienced medicinal chemists. These primarily represent removal of structural alerts or unnecessary functional groups or their replacement with the parent compound structure. Further investigation of tools and objectives to encourage players to perform such removal/replacements themselves is needed. Additionally, as RosettaLigand docking score, which is used internally to evaluate the quality of binding, is not strongly correlated with binding success, further improvement of binding evaluation in-client should increase the rate of player success.

ACKNOWLEDGEMENTS

We dedicate this paper to the memory of Christian Alan Paul Smethurst, without whose tireless support and advocacy this project would not have happened. We thank Alessio Ciulli for consultancy and kindly providing the ^{19}F reporter molecule.

BIBLIOGRAPHY

1 Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756-760, doi:10.1038/nature09304 (2010).

2 Eiben, C. B. *et al.* Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotechnol* **30**, 190-192, doi:10.1038/nbt.2109 (2012).

3 Koepnick, B. *et al.* De novo protein design by citizen scientists. *Nature* **570**, 390-394, doi:10.1038/s41586-019-1274-4 (2019).

4 Khatib, F. *et al.* Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* **18**, 1175-1177, doi:10.1038/nsmb.2119 (2011).

5 Kawrykow, A. *et al.* Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS One* **7**, e31362, doi:10.1371/journal.pone.0031362 (2012).

6 Lintott, C. *et al.* Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies*. *Monthly Notices of the Royal Astronomical Society* **410**, 166-178, doi:10.1111/j.1365-2966.2010.17432.x %J Monthly Notices of the Royal Astronomical Society (2010).

7 Andreasson, J. O. L. *et al.* Crowdsourced RNA design discovers diverse, reversible, efficient, self-contained molecular switches. *Proc Natl Acad Sci U S A* **119**, e2112979119, doi:10.1073/pnas.2112979119 (2022).

8 Soares, P. *et al.* Group-Based Optimization of Potent and Cell-Active Inhibitors of the von Hippel-Lindau (VHL) E3 Ubiquitin Ligase: Structure-Activity Relationships Leading to the Chemical Probe (2S,4R)-1-((S)-2-(1-Cyanocyclopropanecarboxamido)-3,3-dimethylbutanoyl)-4-hydroxy-N-(4-(4-methylthiazol-5-yl)benzyl)pyrrolidine-2-carboxamide (VH298). *J Med Chem* **61**, 599-618, doi:10.1021/acs.jmedchem.7b00675 (2018).

9 Ishida, T. & Ciulli, A. E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. *SLAS Discov* **26**, 484-502, doi:10.1177/2472555220965528 (2021).

10 Zengerle, M., Chan, K. H. & Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem Biol* **10**, 1770-1777, doi:10.1021/acschembio.5b00216 (2015).

11 Mullard, A. Targeted protein degraders crowd into the clinic. *Nat Rev Drug Discov* **20**, 247-250, doi:10.1038/d41573-021-00052-4 (2021).

12 Buckley, D. L. *et al.* Targeting the von Hippel-Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1alpha interaction. *J Am Chem Soc* **134**, 4465-4468, doi:10.1021/ja209924v (2012).

13 Diehl, C. J. & Ciulli, A. Discovery of small molecule ligands for the von Hippel-Lindau (VHL) E3 ligase and their use as inhibitors and PROTAC degraders. *Chem Soc Rev* **51**, 8216-8257, doi:10.1039/d2cs00387b (2022).

14 Loenarz, C. *et al.* Evidence for a stereoelectronic effect in human oxygen sensing. *Angew Chem Int Ed Engl* **48**, 1784-1787, doi:10.1002/anie.200805427 (2009).

15 Kofink, C. *et al.* A selective and orally bioavailable VHL-recruiting PROTAC achieves SMARCA2 degradation in vivo. *Nat Commun* **13**, 5969, doi:10.1038/s41467-022-33430-6 (2022).

16 Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci* **23**, 47-55, doi:10.1002/pro.2389 (2014).

17 Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* **13**, 3031-3048, doi:10.1021/acs.jctc.7b00125 (2017).

18 Tosco, P., Stiefl, N. & Landrum, G. Bringing the MMFF force field to the RDKit: implementation and validation. *J Cheminformatics* **6**, doi:10.1186/s13321-014-0037-3 (2014).

19 Riniker, S. & Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling* **55**, 2562-2574, doi:10.1021/acs.jcim.5b00654 (2015).

20 Ertl, P., Rohde, B. & Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* **43**, 3714-3717, doi:10.1021/jm000942e (2000).
21 Wildman, S. A. & Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences* **39**, 868-873, doi:10.1021/ci990307l (1999).
22 Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform* **1**, 8, doi:10.1186/1758-2946-1-8 (2009).
23 RDKit: Open-source cheminformatics.
24 SeeSAR version 12.1.0 (BioSolveIT GmbH, Sankt Augustin, Germany, 2022).
25 Flare (Cresset Discovery Services, Royston, UK, 2022).
26 Schrödinger (Schrödinger, Inc., 2021-4).
27 Abel, R., Wang, L., Harder, E. D., Berne, B. J. & Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Accounts of Chemical Research* **50**, 1625-1632, doi:10.1021/acs.accounts.7b00083 (2017).
28 Wang, L., Chambers, J. & Abel, R. Protein-Ligand Binding Free Energy Calculations with FEP. *Methods Mol Biol* **2022**, 201-232, doi:10.1007/978-1-4939-9608-7_9 (2019).
29 Alibay, I., Magarkar, A., Seeliger, D. & Biggin, P. C. Evaluating the use of absolute binding free energy in the fragment optimisation process. *Commun Chem* **5**, 105, doi:10.1038/s42004-022-00721-4 (2022).
30 de Castro, G. V. & Ciulli, A. Spy vs. spy: selecting the best reporter for (19)F NMR competition experiments. *Chem Commun (Camb)* **55**, 1482-1485, doi:10.1039/c8cc09790a (2019).
31 Liu, X. *et al.* Discovery of XL01126: A Potent, Fast, Cooperative, Selective, Orally Bioavailable, and Blood-Brain Barrier Penetrant PROTAC Degrader of Leucine-Rich Repeat Kinase 2. *J Am Chem Soc* **144**, 16930-16952, doi:10.1021/jacs.2c05499 (2022).

Table 1

	1 (mix of 4 diastereomers)	**ligand 10 (single enantiomer)**
Affinity/Potency		
NMR K_d [µM]	258 ± 8*	<10
TR-FRET IC_{50} [µM]	264 ± 30	0.090**
Basic Molecular Descriptors		
cLogP	1.21	2.68
Molecular weight	422.5	516.6
H-bond acceptors	5	5
H-bond donors	2	3
TPSA ($Å^2$)	95	111
Rotatable bonds	3	6
Heavy atoms	30	36
Fsp^3	0.27	0.54
***In vitro* ADME Properties**		
Aqueous solubility (pH 6.8) [mg/mL]	>0.098	>0.115
Caco-2 permeability $P_{app, a-b}$ [cm/s] / efflux ratio	$23.0*10^{-6}$/1.4	$2.9*10^{-6}$/15.5

* SD calculated from single titration by averaging results of peaks #1 and #2 from Figure 3

**FP assay as reported previously

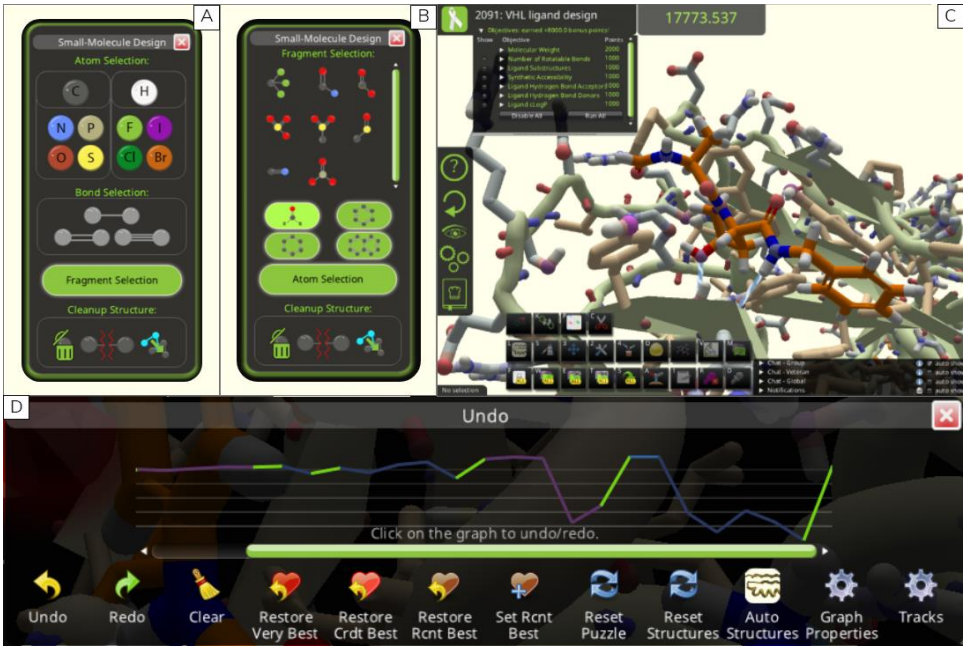

Figure 1: The Drugit user interface. (**a**) Players can use the atomistic design panel to add and delete atoms and bonds. (**b**) The fragment tool allows players to quickly insert functional groups and ring systems. (**b**) Once compounds have been designed, players can optimize their placement with an assortment of tools. Objectives are designed to help guide players by increasing the overall compound score. (**d**) The undo utility function tracks the players progress as they build and manipulate their designed compounds. This function offers a graphical representation of the changes they have made. Each change is denoted by a color associated with the respective change. In the Figure, changes made by the builder are represented in blue, while other changes such as minimization or wiggle are shown in green. Players can use this tool to go back to a state that scored highly and build in different directions.

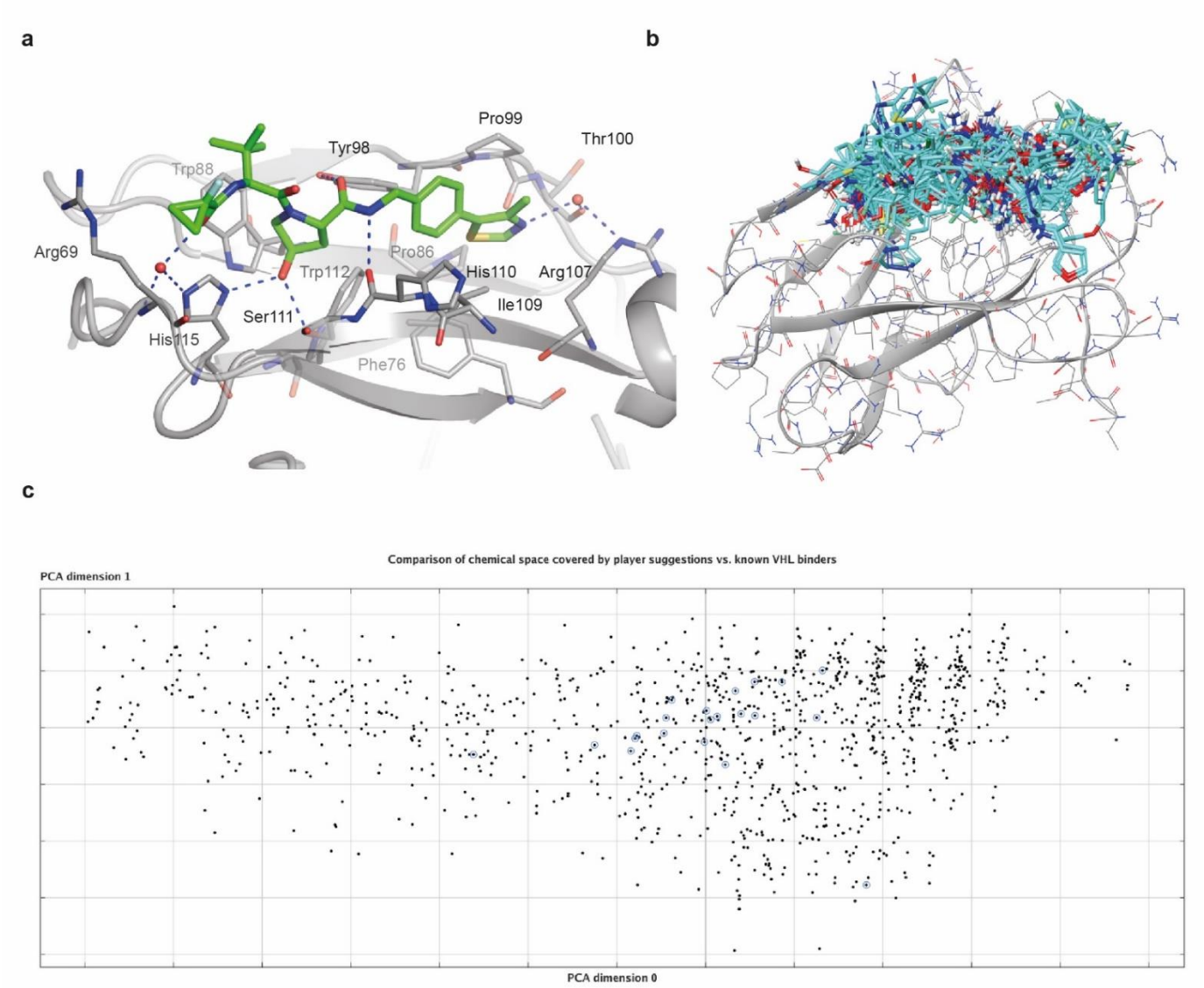

Figure 2: (**a**) Binding mode of the puzzle starting point ligand 10, as observed in the previously reported crystal structure (PDB code: 5NVX; VHL in grey). The ligand 10 is shown as sticks and color coded by atom type (with green C atoms). (**b**) 1073 player suggestions in VHL binding site: The suggestions are displayed as sticks and color-coded by atom type (with cyan C atoms). (**c**) Comparison of the chemical space covered by player suggestions (black dots) vs. known VHL binders used as spike molecules (black dots with blue circles) during property filtering. The chemical space projection is based on the calculation of standard RDkit descriptors, a principal component analysis (PCA), and the reduction of the multidimensional descriptor space to the two dimensions explaining most of the variance.

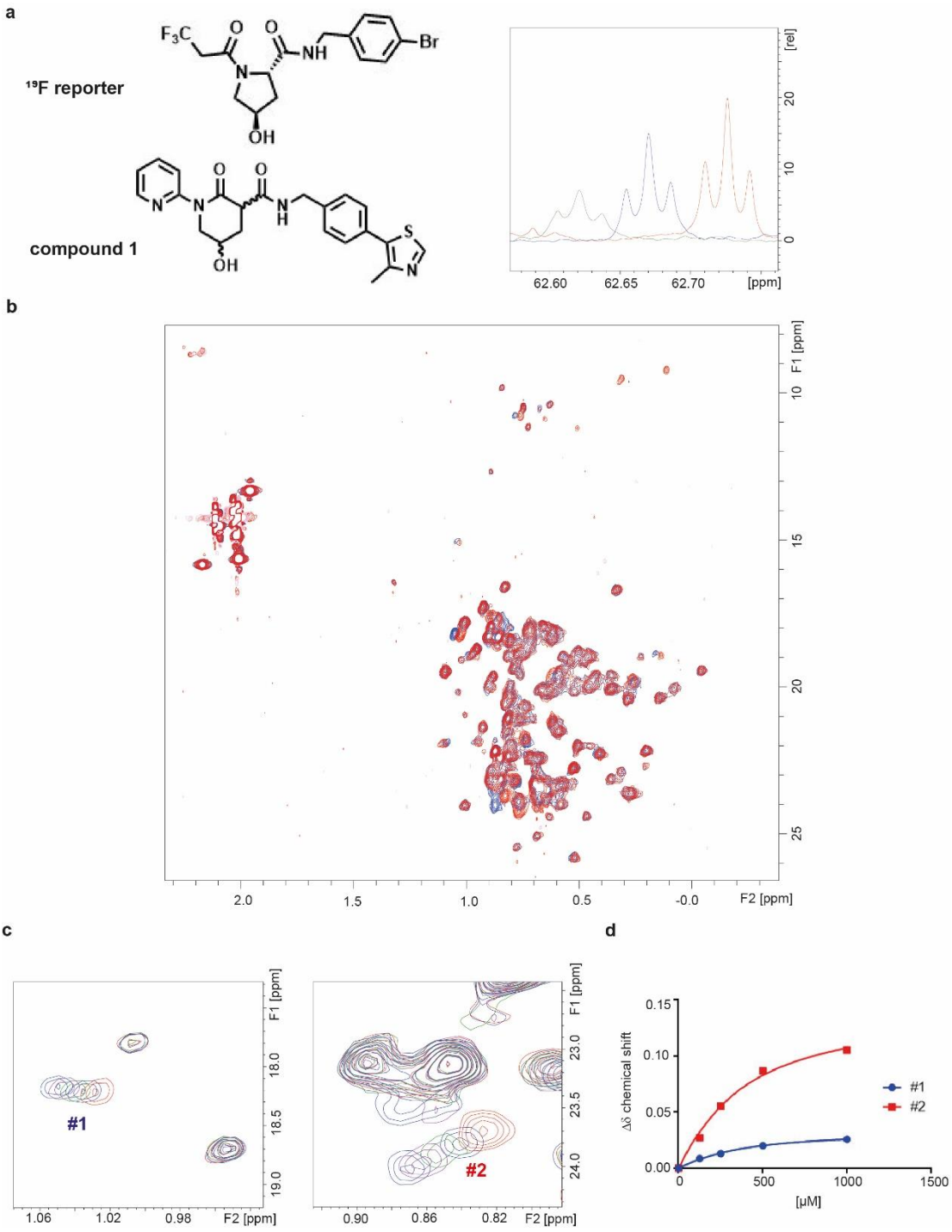

Figure 3: Biophysical triaging. (**a**) Structures of the ^{19}F reporter molecule used to detect binding and of compound **1**. ^{19}F NMR spectrum of 50 µM of the reporter in presence (green) and absence (red) of 2 µM VCB complex. Addition of 500 µM of **1** leads to reappearance of reporter signal due to displacement from the binding site. (**b**) Methyl group region of ^{13}C labeled VCB complex in absence (red) and presence (blue) of 500µM of **1**. (**c**) Expanded regions of signals #1 and #2 monitored during titration of 125µM, 250µM, 500µM and 1000µM of **1**. (**d**) K_D values obtained by fitting the dose-dependent chemical shifts of peaks #1 and #2

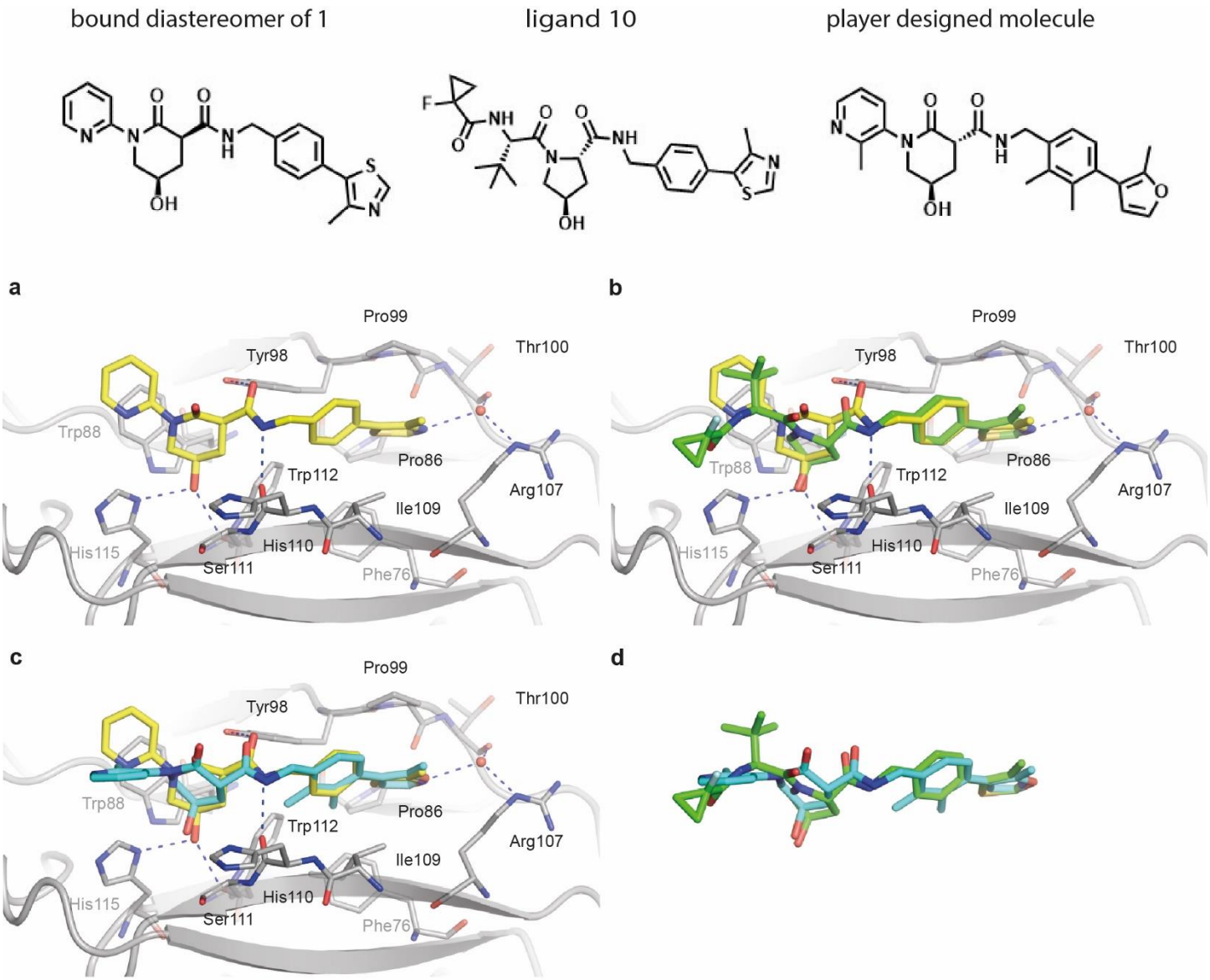

Figure 4: Structural analysis: (**a**) Binding mode of compound **1** as observed in the co-crystal structure with VHL (PDB ID: XXXX), compound **1** is color-coded by atom type with yellow carbons. (**b**) Superposition of the co-crystal structures of compound **1** (PDB ID: XXXX) and ligand 10 in complex with VHL (PDB ID: 5NVX), ligands are color-coded by atom type with yellow and green carbons, respectively (**c**) Superposition of the co-crystal structure of compound **1** and the original player-designed compound complex with the VHL protein coordinates used in the game, ligands are color-coded by atom type and shown with yellow and blue carbons, respectively (**d**) Superposition of the co-crystal structures of ligand 10 in complex with VHL and the player designed compound, ligands are color-coded by atom type with green and blue carbons, respectively, and protein atoms are omitted.