

Pal, S., Nare, Z., Rao, V. A., Smith, B. O., Morrison, I., Fitzgerald, E. A., Scott, A., Bingham, M. J. and Pesnot, T. (2024) Accelerating BRPF1b hit identification with BioPhysical and Active Learning Screening (BioPALS). *ChemMedChem*, 19(6), e202300590. (doi: [10.1002/cmdc.202300590](https://doi.org/10.1002/cmdc.202300590))



Copyright © 2024 Chemistry Europe and Wiley-VCH GmbH.

This is the author accepted manuscript reproduced under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

For the purpose of open access, the author(s) has applied a Creative Commons Attribution license to any Accepted Manuscript version arising.

<https://eprints.gla.ac.uk/321294/>

Deposited on: 5 March 2024

Accelerating BRPF1b hit identification with BioPhysical and Active Learning Screening (*BioPALS*)

Sandeep Pal,^{#[a]} Zandile Nare,^{#[a]} Vincenzo A. Rao,^[a] Brian O. Smith,^[b] Ian Morrison,^[a] Edward A. Fitzgerald,^[c] Andrew Scott,^[a] Matilda J. Bingham^[a] and Thomas Pesnot*^[a]

Abstract: We report the development of BioPhysical and Active Learning Screening (*BioPALS*); a rapid and versatile hit identification protocol combining AI-powered virtual screening with a GCI-driven biophysical confirmation workflow. Its application to the BRPF1b bromodomain afforded a range of novel micromolar binders with

favorable ADMET parameters. In addition to the excellent *in silico/in vitro* confirmation rate demonstrated with BRPF1b, binding kinetics are determined, and binding topologies predicted for all hits. *BioPALS* is a lean, data-rich, and standardized approach to hit identification applicable to wide range of biological targets.

Introduction

Hit identification is arguably one of the most critical stages in the development of new drugs as it defines structural start points for the whole project. Screening campaigns from drug-like collections, fragments, DNA encoded libraries, or virtual databases represent some of the most common strategies employed to hunt for new high-quality hits.^[1,2] In the classical hit identification approach, compound collections are screened using assay cascades tailored to each project and biological target. Whilst this approach aims to provide the most adapted screening technologies for a given biological target, it also requires a wide range of assay technologies and formats to support a drug discovery pipeline, thereby stretching infrastructure costs and project timelines. Instead, a standardized hit identification platform, using a well-defined technology, along with a workflow applicable to most biological targets would provide a new paradigm in hit identification to overcome the aforementioned limitations.

Since its establishment in 2003^[3] over 200,000 biomolecular structures have been deposited on the Protein Data Bank (PDB). This figure is expected to continue its rapid growth thanks to the rise of new techniques such as cryogenic electron microscopy (cryo-EM)^[4] artificial intelligence (AI) such as AlphaFold.^[5] Mirroring the tremendous rise in the number of PDB entries, commercial compound collections such as the REAL database^[6] have grown by several orders of magnitude in recent years. As a result, ultra large-scale structure-enabled virtual screening (e.g., based on docking) is now technically achievable and provides an attractive alternative to classical *in vitro* high throughput screening. Unlike traditional ligand and fingerprint-based virtual screening technologies, brute force docking of vast compound collections requires supercomputers, making it unattainable for many research laboratories. In recent years however, multiple

strategies enabling the virtual screening of large virtual libraries combining docking and AI have been developed,^[7–9] democratizing high scale structure-enabled virtual screening. One of these AI technologies, Molecular Pool-based Active Learning (MolPAL),^[10] was recently developed by Graff and coworkers. MolPAL is underpinned by an Active Learning algorithm which predicts docking scores based on simple molecular fingerprints. The highest scoring compounds are then readily identified from large compound collections despite MolPAL only effectively docking a fraction (typically < 5%) of the overall collection. Although MolPAL has only been applied to a handful of biological targets to date, it provides a lean and versatile structure-based virtual screening technology, which can be deployed across most of the drug discovery landscape.

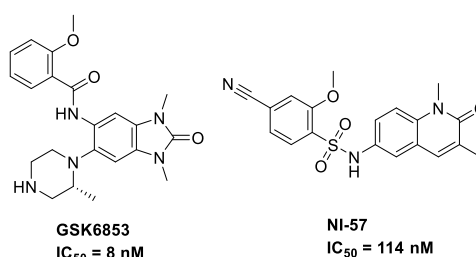


Figure 1: Chemical structures of BRPF1b inhibitors GSK6853^[11] and NI-57^[12] as well as their reported IC₅₀ values for BRPF1b.

A wide range of well-established *in vitro* assay technologies and formats are available to confirm virtual hits^[13–21] (e.g., affinity, activity, phenotypic) but only a few are widely applicable across biological targets. In order to select the most versatile assay technology, we focused on a fundamental aspect of pharmacology, the affinity of a small molecule to its biological target. Biophysical techniques such as surface plasmon resonance (SPR),^[22,23] microscale thermophoresis (MST),^[24,25] and affinity selection-mass spectrometry (AS-MS)^[26] have recently emerged as attractive techniques to measure affinities. Grating-coupled interferometry (GCI) is another biophysical technology enabling fast determination of binding affinities. GCI is applicable to soluble^[27] and membrane bound proteins,^[28–31] RNA targets,^[32] and complex multicomponent assemblies.^[33] GCI also offers the critical advantage of high sensitivity compared to traditional SPR,^[34,35] thus enabling the detection of weak binding events, which are typically associated with primary screening hits.

[a] Dr. S. Pal, Dr. Z. Nare, Dr. V. A. Rao, Mr. I. Morrison, Dr. A. Scott, Dr. M. J. Bingham, Dr. T. Pesnot
Concept Life Sciences
Frith Knoll Road, Chapel-en-le-Frith, High Peak, SK23 0PG, UK
*E-mail: tom.pesnot@conceptlifesciences.com

[b] Dr. B. O. Smith
University of Glasgow
School of Molecular Biosciences, College of Medical Veterinary and Life Sciences, G12 8QQ, UK

[c] Dr. E. A. Fitzgerald
Creoptix AG
Zugerstrasse 76, CH-8820 Wädenswil, Switzerland

Authors contributed equally to this work.

This high sensitivity is maximized by the waveRAPID® (Repeated Analyte Pulses of Increasing Duration) technology,^[36] which accelerates the acquisition of binding kinetic parameters by screening compounds at a single concentration and varying the injection time on the sensor chip. This is compared to traditional SPR where compounds are typically screened using the same injection time for a concentration series.

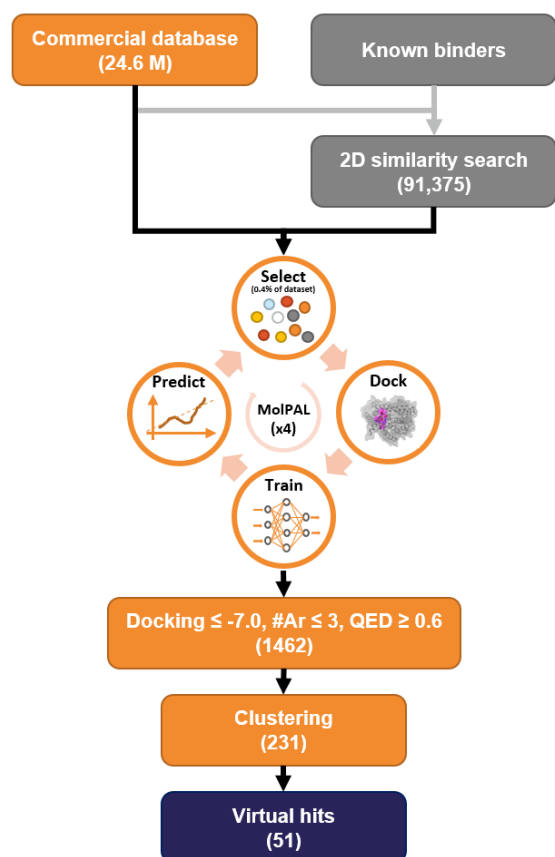


Figure 2: Virtual screening piece of the **BioPALS** protocol, centred on the MoIPAL algorithm. Activities in orange are integral to the protocol, activities in grey can be added depending on the drug discovery strategy and data available on the target. Figures in bracket shows the number of compounds processed through each individual stage for the BRPF1b bromodomain.

The bromodomain and PHD finger-containing protein 1b (BRPF1b) protein-protein interaction module contains both epigenetic acetyl reader and scaffolding functions and was recently identified as a therapeutic target to treat hepatocellular carcinoma^[37] and acute myeloid leukemia,^[38] an aggressive family of cancers with a 5-year survival rate below 30% for adults.^[39] A range of BRPF1b bromodomain inhibitors have been reported, including **GSK6853**,^[11] and **NI-57**^[12] (**Figure 1**). BRPF1b inhibitors were identified through a wide range of hit identification technologies including high throughput fragment docking,^[40] ligand-based screening,^[41] and hit optimization guided by numerous X-ray ligand-protein structures.^[11,40–42] Despite these efforts, a selective BRPF1b inhibitor has yet to enter clinical

investigation and new chemotypes inhibiting this target are still highly sought after.

Here, we report the development of the Biophysical and Active Learning Screening (**BioPALS**) hit identification protocol, and its application in the identification of novel BRPF1b lead compounds. The initial *in silico* stage of the protocol involves a MoIPAL-driven virtual screen and provided a selection of 51 virtual hits from a commercial collection of over 24 million compounds. *In vitro* confirmation of the virtual hits (using GCI) highlighted 36 primary BRPF1b binders, including 20 hits with affinities < 250 μ M. Importantly, these hits are selective for BRPF1b over its inactive isoform, BRPF1a.^[42] Orthogonal confirmation of target engagement (by differential scanning fluorimetry (DSF) and ligand observed nuclear magnetic resonance (NMR)) and ADMET profiling, highlighted several attractive leads for subsequent development of novel BRPF1b chemotypes. In addition to the excellent *in silico/in vitro* hit confirmation rate exemplified in this study, **BioPALS** provides high-quality hits with fully characterized binding kinetics and predicted binding topologies, thus offering an accelerated and versatile path to aid drug discovery.

Results

BRPF1b is a well-established oncology target for which inhibitors belonging to many different chemical series have been identified. These ligands provide an opportunity to focus the MoIPAL protocol on chemical matter structurally related to established BRPF1b inhibitors. We therefore initiated the virtual screening piece of the **BioPALS** protocol by carrying out a similarity search on 27 known BRPF1b inhibitors with $pIC_{50} > 7.2$ (structures provided in SI **Figure 1**) against a commercial library of over 24 million compounds (eMolecules collection). 91,375 structures were identified, providing a pool of commercial compounds suitable for BRPF1b MoIPAL screening (**Figure 2**, grey). In instances where there is no known ligand for the target of interest or, if an unbiased sampling of the chemical universe is preferred, this step can be excluded from the **BioPALS** protocol.

Next, we submitted the 91,375 commercial analogues to the MoIPAL docking protocol using PDB ID: 6EQK,^[41] a high-resolution crystal structure of BRPF1b. An initial benchmarking was carried out using a 0.1% random sampling and four rounds of active learning cycles (combination of random forest (RF) and directed message passing neural network (D-MPNN) machine learning models with greedy and upper confidence bound (UCB) acquisitions) to select compounds with the best docking scores, while still providing a structurally diverse set. As shown in **Figure 3A**, after four MoIPAL cycles 85% of the 368 compounds selected by the algorithm (blue) display a docking score below < -7.0, compared to only 50% for a random selection of the 91,375 similarity hits (grey). In addition, the average docking score shifted from -7.0 for the random selection to -8.0 for the MoIPAL selection, a significant improvement considering the random selection is populated with compounds already structurally similar to potent BRPF1b inhibitors. Hence, the MoIPAL algorithm efficiently

selects compounds with better docking score despite only docking 0.4% of the overall dataset. This recapitulates the performance reported in seminal MolPAL studies,^[10] and further demonstrates that the algorithm provides a lean technology for structure-enabled virtual screening. In addition, compounds selected by the *in silico* piece of the **BioPALS** protocol highlighted virtual BRPF1b hits (blue) complying with the Rule of 5^[43] (orange) and displaying physicochemical profiles suitable for lead generation (**Figure 3B**).

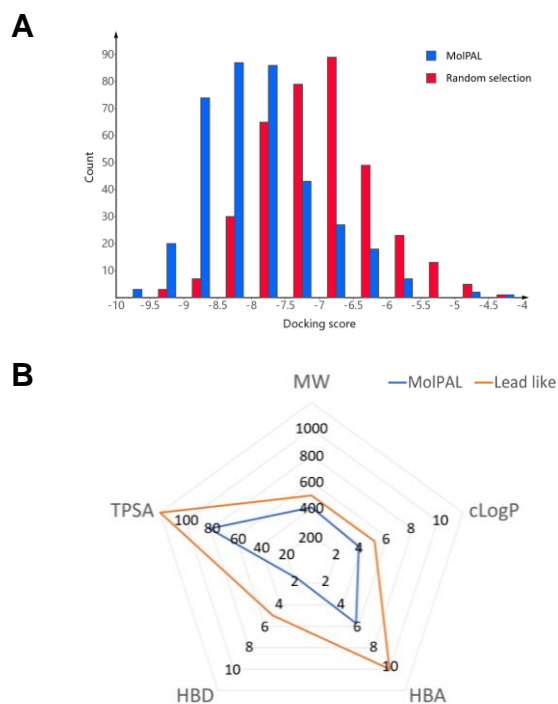


Figure 3: Benchmarking the MolPAL algorithm. **(A)** Docking of a random selection shows normal distribution with mean around -7.0 (red), compared to an average docking score of -8.0 following MolPAL selection (red). **(B)** Physicochemical properties of MolPAL hits (blue) compared to lead-like boundaries (orange). MW (g.mol⁻¹) = Molecular weight, HBA = number of hydrogen bond acceptors, HBD = number of hydrogen bond donors, TPSA (Å²) = topological surface area.

With the protocol successfully benchmarked, we proceeded to apply MolPAL with a 0.4% sampling over four active learning cycles, to increase the number of virtual hits prior to downstream triaging. The resulting 1462 compounds identified by the algorithm were triaged by docking score (≤ 7), number of aromatic rings (≤ 3), and quantitative estimate of druglikeness ($\text{QED} \geq 0.6$). A final clustering step using Morgan circular fingerprints completes the *in silico* piece of the **BioPALS** protocol and afforded 51 commercially available virtual hits with BRPF1b binding topologies predicted by MolPAL docking.

The virtual screening piece of the **BioPALS** protocol provides a lean strategy to identify new chemical matter predicted to bind the biological target of interest at a precisely defined site. The process

can be run on any target for which a structure has been deposited or predicted, and a binding site is defined. Whilst **BioPALS** is underpinned and accelerated by the structure-enabled MolPAL virtual screening algorithm, additional cheminformatic tools (e.g., 2D similarity searches, CNS-MPO filtration) can be embedded within the protocol if relevant to the biological target classes or therapeutic indications.

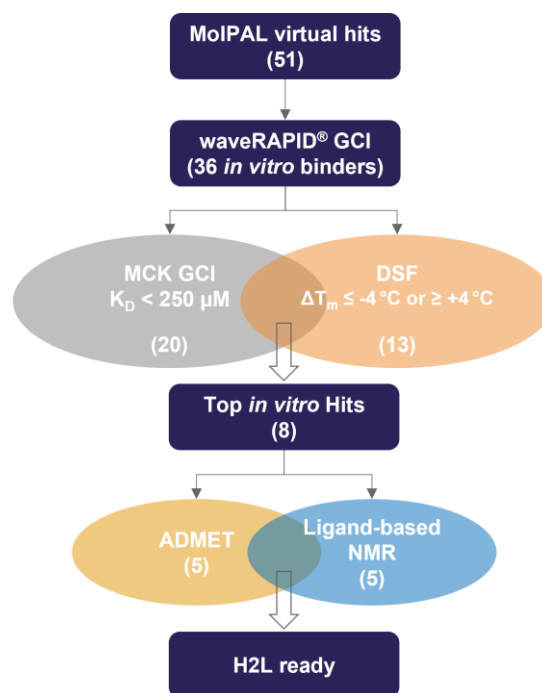


Figure 4: **BioPALS** *in vitro* screening confirmation workflow including GCI, DSF, ligand observed NMR and ADMET profiling. Figures in brackets correspond to the number compounds taken through each step of the protocol.

We next concentrated on the *in vitro* confirmation of the virtual hits, as well as their profiling and triaging, prior to initiating hit-to-lead development. Akin to the virtual screening piece of the **BioPALS** protocol, the main aims of *in vitro* confirmation are to identify high-quality hits, in a lean and systematic manner applicable across target classes and therapeutic areas. To satisfy these criteria, we opted for target engagement (i.e., affinity-based) assays as they arguably measure the most fundamental feature driving a pharmacological event, whatever the biological target of interest. We selected three biophysical techniques (DSF, ligand observed NMR, and GCI) to enable the orthogonal confirmation of the obtained binding data. These three technologies were also carefully selected as they each provide complementary information, aiding hit prioritization and optimization for all subsequent steps of the discovery process. GCI accurately determines binding affinities, kinetics, and residence time, even for weak binders typically identified in the early stages of the drug discovery process. DSF characterizes the impact of each ligand on the thermal stability of the target protein. Finally, ligand observed NMR provides *in vitro* evidence of the binding

topologies predicted by the MolPAL algorithm. Hence, by combining GCI, with DSF and NMR, we aim to characterize binding affinities, kinetics, and topologies in a single and generally applicable workflow (**Figure 4**).

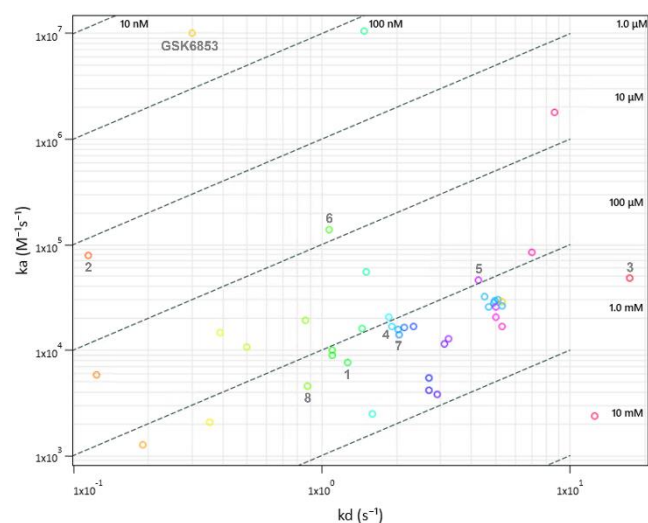


Figure 5: Two-dimensional isoaffinity kinetic plot of association (k_a) and dissociation (k_d) rate constants from a waveRAPID[®] screen of 51 virtual hits. The 36 primary hit binders are shown. Diagonal lines indicate equilibrium binding constants (K_D) and are shown to visualize affinity distribution. Each circle represents a binder and is coloured according to the dissociation rate. The top eight hits (**Table 1**) and the **GSK6853** positive control compound are highlighted in grey text.

All 51 virtual hits (100 μ M) were first tested against BRPF1b using the waveRAPID[®] GCI screening technology to eliminate non-binders. Thirty-six primary BRPF1b binders with K_D values ranging from 0.14 μ M to 1 mM were identified (**Figure 5**). Hits were selected based on sensorgrams showing clear concentration dependent binding ($R_{max} > 2.0$ pg/mm²) in the dedicated WAVEcontrol software (version 4.5.17; Creoptix AG, Switzerland) with low statistical association/dissociation rate errors < 60 , and kinetic rate constants within the measurable limits (i.e., $k_a = 10^3 - 5 \times 10^7$ M/s⁻¹; $k_d = 10^{-6} - 10$ s⁻¹). Affinity was not used in the hit selection criteria because the focus of this work was to confirm the suitability of the **BioPALS** workflow for assessing *in vitro* target engagement of virtual hits, rather than identification of high affinity binders. Moreover, the hits identified in this workflow may serve as good lead molecules in instances where there is scope to improve potency. Most waveRAPID[®] hits displayed moderate to fast binding kinetics, including fast dissociation rates. Two BRPF1b binders displaying K_D values < 1 μ M in the primary screen (one of which is not shown in **Figure 5** for clarity) were false positives whose binding could not subsequently be orthogonally confirmed due to solubility issues. Interestingly, the higher affinity observed for the **GSK6853** positive control over the virtual hits is primarily due to fast association rather than slow dissociation times. This kinetic profile implies that **GSK6853** has a low residence time despite having nanomolar affinity for BRPF1b and may set more stringent requirements for DMPK properties, in particular the duration of

exposure at the site of action. This result, combined with our kinetics driven **BioPALS** protocol, highlights new opportunities to develop novel potent BRPF1b inhibitors with enhanced target residence time.

We confirmed and prioritized the waveRAPID[®] hits using a combination of GCI multi-cycle binding kinetics (MCK) and DSF. MCK experiments confirmed 29 of the 36 primary hit binders including 20 with K_D values < 250 μ M, and 10 with $K_D < 100$ μ M. This suggests that the medium throughput waveRAPID[®] technology is suitable to triage compounds, but MCK is required to ascertain binding affinities and eliminate false positives. The same GCI workflow was applied to BRPF1a to determine selectivity of the 51 virtual hits. Initial waveRAPID[®] data identified two binders ($K_D = 300$ μ M and 700 μ M) with only one being confirmed in MCK experiments. Importantly, these two BRPF1a binders were deselected early in the BRPF1b workflow. Taken together, these data confirm that the **BioPALS** protocol can identify selective binders, in this case for BRPF1b.

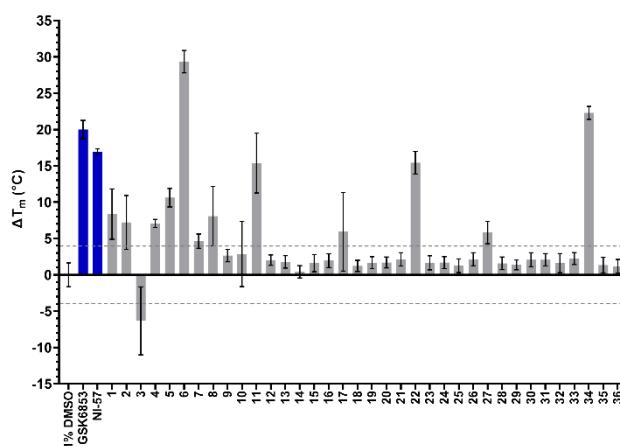


Figure 6: DSF characterization of the 36 waveRAPID[®] primary hits (grey). Known BRPF1b inhibitors (**GSK6853** and **NI-57**) are coloured in blue, and the $> +4$ $^{\circ}$ C and < -4 $^{\circ}$ C thresholds are indicated with grey dashed lines.

Alongside MCK, we confirmed the 36 primary BRPF1b waveRAPID[®] hits using an orthogonal DSF screen with compounds at 100 μ M (**Figure 6**). The assay was first validated with **GSK6853** and **NI-57**, for which thermal shifts of $+20$ $^{\circ}$ C and $+17$ $^{\circ}$ C were observed, respectively. Importantly, small ΔT_m shifts (e.g., $\pm 1 - 2$ $^{\circ}$ C) can be indicative of false positive hits or very weak interactions. Therefore, to limit this risk, we set higher ΔT_m thresholds of < -4 $^{\circ}$ C and $> +4$ $^{\circ}$ C. 13 of the 36 primary hits were shown to interact with BRPF1b either by stabilizing ($\Delta T_m > +4$ $^{\circ}$ C) or destabilizing (ΔT_m decreased by over < -4 $^{\circ}$ C), including five compounds inducing a ΔT_m over $+10$ $^{\circ}$ C. Although no correlation was observed between the K_D values determined by GCI and ΔT_m values determined by DSF, the orthogonality of the two biophysical technologies provides an additional means to eliminate potential false positives. The eight hits displaying a $K_D < 250$ μ M (determined by GCI MCK) and a thermal shift $> +4$ $^{\circ}$ C or < -4 $^{\circ}$ C (displayed in **Table 1**) were prioritized as *in vitro* hits.

These structures are all lead-like and diverse thanks to the parametrization of MolPAL and additional cheminformatics filtration embedded within the *in silico* piece of **BioPALS**. It is worth noting the protocol highlighted compounds **4** and **5**, (**Table 1**) which have scaffolds related to **NI-57** and **GSK6853**, respectively. Whilst this suggests **BioPALS** can identify developable hits, it might also be associated with the initial similarity search of the protocol. Hence, removing the initial similarity search step might be favorable when identifying novel chemical matter is critical to the project.

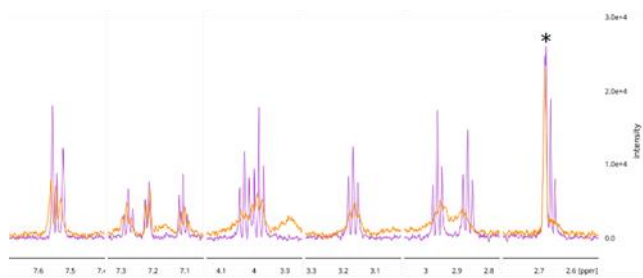


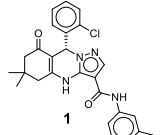
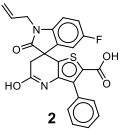
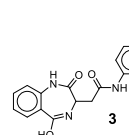
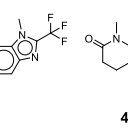
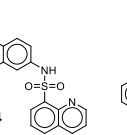
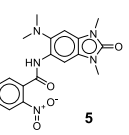
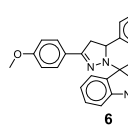
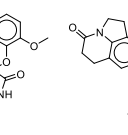
Figure 7: Sections of the 1D ^1H NMR spectra illustrating peaks from compound **7** (200 μM) in the absence (pink) and presence (orange) of the BRPF1b bromodomain (26 μM). The asterisk denotes a peak from a contaminant component of the sample.

A structurally diverse set of five compounds (**2**, **3**, **4**, **5**, and **7**; **Table 1**) were further prioritized based on their overall biophysical profile and progressed for ligand observed NMR characterization and *in vitro* ADMET profiling. One-dimensional (1D),^[44] saturation-transfer difference (STD),^[45] water-ligand observed via gradient spectroscopy (waterLOGSY)^[46] and Carr-Purcell-Meiboom-Gill (CPMG)^[47] NMR spectra were acquired for each ligand (200 μM) in the presence or absence of BRPF1b protein (26 μM). Significant signal differences were observed in one or more of the NMR experiments for compounds **2**, **3**, **5**, and **7** (example shown in **Figure 7**), whereas compound **4** was insufficiently soluble to be validated using this technique. Altogether, ligand observed NMR provides yet further evidence that the indicated hit compounds

identified in the **BioPALS** protocol are bona fide BRPF1b binders. In addition, for compound **5** the relative magnitude of the chemical shift changes observed upon addition of the protein suggest that the benzimidazolone methyls are most intimately involved in binding (data not shown), recapitulating data reported by Bamborough and coworkers, using X-ray crystallography.^[11] For all other hits however, the size of the recombinant BRPF1b bromodomain (15 kDa) proved too restrictive to provide any tangible evidence of specific protons being involved in binding to BRPF1b.

This limitation is however compensated by MolPAL which predicts binding topologies, hence providing modelling hypotheses which can readily be tested experimentally (e.g., through the profiling of hit analogues in targeted hit exploration studies). This is exemplified for BRPF1b in **Figure 8** where the lowest energy binding topologies predicted by MolPAL (as part of the **BioPALS** protocol) are shown for the five prioritized hits **2**, **3**, **4**, **5**, and **7**. All molecules are buried within the BRPF1b active site and interact with Phe714 (through face-to-face π - π interactions) as reported in many other BRPF1b structures such as those of **GSK6853** (PDB ID: 5G4R)^[11] and **NI-57** (PDB ID: 5MYG).^[12] Several other protein-ligand interactions including H-bonding to the backbone of Asp651, Ile652, Glu655, and the amide side chain of Asn708 were highlighted, some of which have also been previously reported.^[41] It is also worth noting that the binding topology of compound **4** shown in **Figure 8C** resembles that of its structurally related analogue, **NI-57** previously determined by X-ray crystallography with the (dihydro)quinolone making direct π -stacking interaction with Phe714. In contrast, the lowest energy docking pose of compound **5** predicted by the MolPAL algorithm differs from that of its related analogue **GSK6853** (previously determined by X-ray crystallography). In our model, the Phe714 residue interacts with the nitro-phenyl moiety (absent in **GSK6853**), whilst it interacts with the dimethylbenzimidazolone in the **GSK6853** X-ray structure. This discrepancy was highlighted in the ligand observed NMR assay, which suggests that compound **5** binds in a topology akin to the **GSK6853** X-ray structure, confirming the importance of including ligand observed NMR in the protocol.

Table 1: Structures and completed biophysical profiling of the eight prioritized BRPF1b hits selected from the **BioPALS** protocol.

Compound	MW (g.mol ⁻¹)	K_D (μM)	k_a (M/s ⁻¹)	k_d (s ⁻¹)	ΔT_m (°C)	NMR binding
	461	27.7*	1.7×10^4	0.5	$+ 8.4 \pm 3.5$	n.d.
	448	43.6 ± 2.2	4.6×10^4	2.0	$+ 7.2 \pm 3.7$	Yes
	431	46.5 ± 5.6	4.8×10^4	2.2	$- 6.3 \pm 4.7$	Yes
	367	61.2 ± 14.9	3.5×10^4	2.0	$+ 7.1 \pm 0.6$	Insoluble
	369	128.1 ± 87.0	3.0×10^4	2.1	$+ 10.6 \pm 1.3$	Yes
	427	133.7*	5.8×10^3	0.8	$+ 29.3 \pm 1.5$	n.d.
	354	186.2 ± 11.1	8.4×10^3	1.6	$+ 4.6 \pm 1.0$	Yes
	437	204.9 ± 30.3	6.1×10^3	1.2	$+ 8.1 \pm 4.1$	n.d.

K_D , k_a and k_d were generated using GCI MCK, ΔT_m generated by DSF, and ligand observed NMR binding determined as an overall assessment from 1D, STD, waterLOGSY, and CPMG experiments. * = data shown for single replicate only, n.d. = not determined.

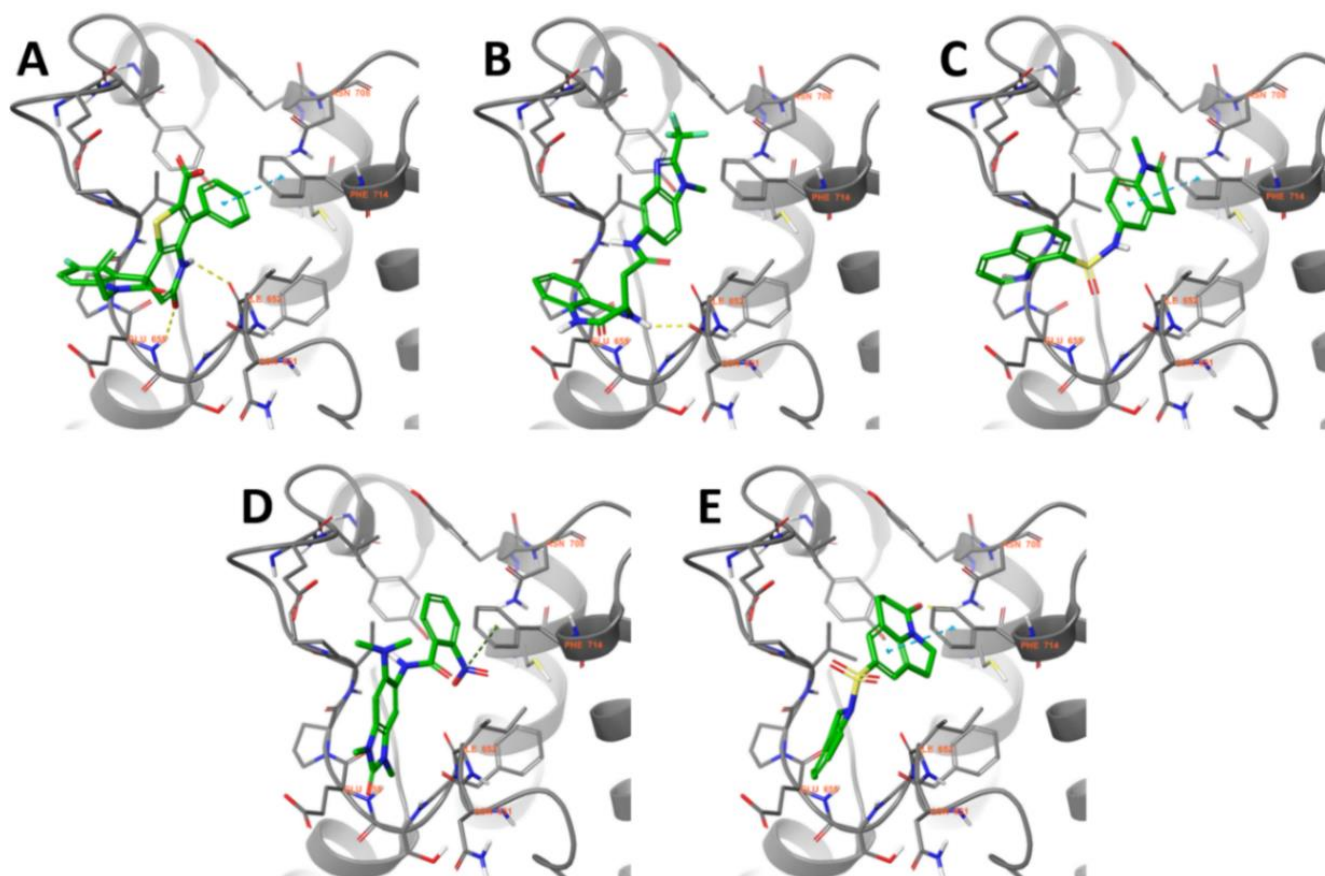


Figure 8: Lowest energy docking poses predicted by MolPAL for compounds **2**, **3**, **4**, **5** and **7** (Figures **A**, **B**, **C**, **D** and **E**, respectively; green sticks) within the BRPF1b active site (grey cartoon). Key residues are represented in grey sticks and amino acid sequence numbers highlighted with orange labels. Nitrogen, Oxygen and Sulfur atoms are coloured blue, red and yellow respectively. Key protein-ligand interactions are represented with dotted lines (hydrogen bonds in yellow, π - π interactions in cyan and π -cation interactions in green).

The *in vitro* ADMET profiling of the five prioritized hits was performed as the final step of the **BioPALS** protocol to assess their developability as lead series. As shown in **Table 2**, all compounds, except **7**, display low *in vitro* clearance in human microsomes and hepatocytes. Free fraction from plasma ranges from 5 to 23%, including for the carboxylic acid containing compound **2**. Additionally, CYP450 inhibition profiling of compounds **2** and **3** confirms IC_{50} values for all five main isoforms (1A2, 2C9, 2C19, 2D6 or 3A4) were above 10 μ M, thus de-risking

drug-drug interaction liability for these hits. Altogether, compounds **2**, **3**, **4** and **5** stand out as promising BRPF1b hits. Compounds **2** and **3** are of particular interest since these are novel chemotypes shown to selectively bind BRPF1b over BRPF1a and displaying promising *in vitro* ADMET profiling. Their structure affinity relationship (SAR) optimization will not only be facilitated by the in-depth biophysical profiling provided by the *in vitro* piece of the **BioPALS** protocol, but also by the binding topologies predicted by the MolPAL algorithm.

Table 2: <i>In vitro</i> ADMET profile of the five prioritized BRPF1b hits.						
Compound	MPPB (Fu)	MLM Clint (μ L.min ⁻¹ .mg ⁻¹)	RLM CLINT (μ L.min ⁻¹ .mg ⁻¹)	HLM Clint (μ L.min ⁻¹ .mg ⁻¹)	HHeps Clint (μ L.min ⁻¹ .10 ⁶ cells)	CYP450 inhibition (μ M)
2	0.18	17	41	< 5	< 3	All > 50
3	0.12	< 5	< 5	< 5	< 3	All > 10
4	0.23	20	89	10	4	n.d.
5	0.12	6	< 5	< 5	< 3	n.d.
7	0.05	157	216	94	67	n.d.

CYP450 inhibition was measured as a pooled mixture of 1A2, 2C9, 2C19, 2D6 and 3A4 isoforms. MPPB = mouse plasma protein binding, MLM = mouse liver microsomal stability, RLM = rat liver microsomal stability, HLM = human liver microsomal stability, HHeps = human hepatocyte stability, Fu = fraction unbound, n.d. = not determined.

Discussion

Whether the aim is to deorphanize a novel biological target, find a backup chemical series, or circumvent a challenging IP space, hit identification is a necessary step for the development of new drugs. In this work we present **BioPALS**, a workflow leveraging cutting-edge machine learning and biophysical technologies to provide a systematic hit identification solution applicable to a wide range of biological targets.

Virtual screening is a well-established hit identification strategy, yet historically hindered by high CPU requirements and the simplistic (typically ligand-based) models used. In recent years however, thanks to the significant strides made in structural biology and AI, structures (X-ray and Cryo-EM) and models (AlphaFold) are now available for most proteins. Moreover, high throughput docking against these biological targets is achievable even with standard computers. MolPAL is an example of emerging technologies crystallizing these advances, and potentially offering a generally applicable technique for hit identification. When applying MolPAL to the BRPF1b bromodomain (following an initial similarity-based triaging), 51 virtual hits were selected from a library of over 24 million commercial compounds; the overall *in silico* process being completed in a matter of days.

In the subsequent *in vitro* stage of the **BioPALS** workflow we combined three complementary biophysical techniques, GCI, DSF, and ligand observed NMR to confirm and prioritize the MolPAL hits. Akin to MolPAL, all three techniques focus on binding affinities and direct target engagement. GCI was selected for the initial screening process as it is a highly sensitive technology able to determine affinities from the picomolar to the millimolar range, and applicable to soluble and membrane bound proteins as well as complex ternary structures. *In vitro* screening of the 51 virtual hits using GCI waveRAPID[®] afforded 36 primary binders, including 20 with $K_D < 250 \mu\text{M}$ (confirmed by GCI MCK). The resulting 39% hit rate suggests the virtual screening step of the **BioPALS** workflow can efficiently select binders from large virtual collections. A parallel counter-screen of the 51 virtual hits against BRPF1a (inactive isoform of BRPF1b) did not yield any hits with $K_D < 250 \mu\text{M}$, suggesting that all hits are selective for the BRPF1b Isoform. Although not discussed for BRPF1b, GCI was also selected for its high sensitivity in determining binding kinetics, one of the key parameters monitored to drive the hit optimization process. While further hit validation in cell-based assays would be the logical next step for future studies, we did not deem it critical to demonstrate the utility of **BioPALS** in this work because target engagement was confirmed using biophysical methods.

Biophysical technologies can suffer from high false positive rates and confirming binders using orthogonal techniques is best practice. We selected DSF as the orthogonal confirmation screen. This technique measures the effect of ligands on the thermal stability of the biological target of interest, thus not only providing an orthogonal confirmation of target engagement but also complementary data to binding kinetics generated by GCI. We

confirmed that 13 of the 36 primary GCI hits cause a significant shift in the thermal stability of BRPF1b ($> +4 \text{ }^\circ\text{C}$ or $< -4 \text{ }^\circ\text{C}$). Combining both DSF and GCI led to the identification of eight validated hits, giving a 16% orthogonally confirmed hit rate.

Ligand observed NMR was selected as the third and final orthogonal technique to further confirm target engagement and provide *in vitro* evidence of the binding topologies predicted by MolPAL without the need to acquire new ligand-protein structures. Five out of the eight confirmed hits were tested in 1D, waterLOGSY, STD, and CPMG NMR studies. Binding was confirmed for all compounds sufficiently soluble at the high concentration required by the assay, providing yet more evidence of the MolPAL hits binding BRPF1b. Ligand observed NMR may be one of the main restrictions associated with the **BioPALS** workflow as it is only applicable to soluble proteins large enough to have a significantly slower tumbling rate than its cognate ligand. Although BRPF1b is a soluble protein, its small size (15 kDa), meant protons involved in ligand binding to BRPF1b could not be confirmed with high confidence. For these proteins and others unsuitable for ligand observed NMR studies, we can however use **BioPALS** predictions to drive the subsequent hit-to-lead process.

Besides pharmacology (i.e., binding affinity, kinetics, and topology), the quality of screening hits is also defined by its ADMET profile. We therefore completed the **BioPALS** workflow by determining the microsomal and hepatocyte stability as well as other parameters deemed critical for the targeted therapeutic intervention, in this instance CYP450 inhibition to reduce the risk of drug-drug interactions for poly-medicated cancer patients. This led to the identification of compounds **2** and **3**, both novel and selective BRPF1b binders, displaying promising *in vitro* ADMET parameters.

Conclusion

Herein, we report the development of **BioPALS**; an innovative hit identification workflow leveraging recent advances in artificial intelligence and biophysics to identify and characterize high-quality hits. Its application to the BRPF1b bromodomain afforded several micromolar binders with excellent *in vitro* ADMET parameters. The **BioPALS** workflow focuses on direct target engagement (via docking and binding affinity), the most fundamental driver of pharmacology and aims to be readily applicable to most biological targets. Although some limitations are present, including the requirement for a defined binding site, and MolPAL's ability to detect binders eliciting the desired pharmacology for highly dynamic targets (e.g., some GPCRs and ion channels), these can be overcome with simple additional steps to the workflow. These include *in silico* approaches such as molecular dynamics and active site scouting, as well as the *in vitro* determination of activity, potency, or efficacy in additional bioassays. Despite these limitations, **BioPALS** offers a standardized hit identification process suitable to most biological targets and business models. With the continuous improvements

in AI algorithms and biophysics for drug discovery, we expect this type of approach will be increasingly utilized in years to come.

Experimental section

Compound library preparation: A list of 27 BRPF1b active inhibitors with an IC_{50} of $\leq 1 \mu\text{M}$ were downloaded from ChemBL^[48] (**SI Figure 1**). Next, 2D similarity searches on these active inhibitors were performed on eMolecule's screening collection^[49] of 24.6 million compounds using a Tanimoto^[50] cut-off of 0.5. The resulting library of hits contained 91,375 compounds. MolPAL was applied to this library to rapidly identify compounds that docked best into the BRPF1b active site.

BRPF1b docking: A high-resolution co-crystal structure of human BRPF1b bromodomain protein (PDB ID: 6EKQ) was selected.^[41] Structure preparation was carried out using the protein-preparation wizard in Maestro.^[51] The ligand binding site in the crystal structure was targeted and used as the centroid of the grid box for docking studies. Water molecules in the active site were removed.^[52,53]

Virtual screening workflow: The virtual screening workflow used in this study is shown in **Figure 2**. The docking process, embedded within the MolPAL workflow was carried out using AutoDock Vina 1.1.2 and PDB ID: 6EKQ, as stated above. We used RF and D-MPNN machine learning models in combination with greedy and UCB acquisition criteria as the main MolPAL parameters for our studies. The method randomly chose 1% of the master library at the start and for all subsequent acquisition cycles. In four active learning cycles a total of 1462 compounds were highlighted. Compounds from the four screens (RF+greedy, RF+UCB, D-MPNN+greedy, and D-MPNN+UCB) were annotated and combined into one single file for further filtering. Physicochemical properties and druglikeness^[54] of the compounds were calculated using RDKit.^[55] Clustering was performed on the compound library using Morgan circular fingerprints^[56] with a radius of two, a bit-vector length of 2048, and a Tanimoto cut-off of 0.6. Unique compounds from a cluster with the highest druglikeness scores were selected. The resulting library was further restricted to 320 compounds by limiting the maximum number of aromatic rings to two. Finally, 51 compounds were selected based on cost, lead time and synthetic tractability, and their potential binding assessed using bio-physical methods.

Sourcing of virtual hits: Virtual hits were sourced from eMolecules.^[49] The structural integrity of each compound was confirmed by ultra-high performance liquid chromatography-mass spectrometry (UPLCMS) analysis prior to *in vitro* testing. Compounds **2**, **3**, **4**, **5** and **7** were determined to be > 95% pure by LCMS analysis, (**SI Figure 3-7**).

Recombinant protein production and purification: A truncated version of the BRPF1b gene (UniProt: P55201), comprising solely of the bromodomain (residues 622 – 738) was codon optimized for optimal expression in *Escherichia coli* and

synthesized (GenScript). The resulting insert was sub-cloned into NdeI and XhoI cleaved pET28(+)-TEV vector (GenScript) for protein production. The resulting plasmid, BRPF1b_pET28(+)-TEV, encodes an N-terminal hexa-histidine tag followed by a Tobacco Etch Virus (TEV) protease recognition site. Protein expression in *E. coli* BL21 (DE3) cells, and purification were performed by the Edinburgh Protein Production Facility (EPPF) according to published methods.^[11] A yield of $\sim 6 \text{ mg L}^{-1}$ of culture was achieved and BRPF1b purity > 85% was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). BRPF1b was eluted from the size exclusion column using 50 mM HEPES pH 7.4, 300 mM NaCl, 0.1 mM DTT before concentrating to 2 mg mL^{-1} using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore). BRPF1b was subsequently aliquoted, flash frozen and stored at $-80 \text{ }^\circ\text{C}$ for further use. Recombinant BRPF1a (ab196416) was purchased from Abcam (Cambridge, UK).

DSF: DSF^[57] was performed using the Protein Thermal Shift™ Dye Kit (Thermo Fisher Scientific) to determine the effect of virtual hits and control compounds (**GSK6853** and **NI-57**) on thermal stability of recombinant BRPF1b. Solutions of BRPF1b in 50 mM HEPES pH 7.4, 300 mM NaCl, 0.1 mM DTT were prepared in Protein Thermal Shift™ Buffer and $1 \times$ Thermal Shift™ Dye. Final concentrations of BRPF1b and compounds were $7.5 \mu\text{M}$ and $100 \mu\text{M}$ (1% (v/v) DMSO), respectively. Reactions (20 μL) were prepared in triplicate on ice in a 384-well thin-wall PCR plate (Thermo Fisher Scientific) and immediately sealed with an adhesive PCR seal (Bio-Rad). Following centrifugation (2 mins, $200 \times g$, $4 \text{ }^\circ\text{C}$) fluorescence was monitored in a QuantStudio5 qPCR machine (Thermo Fisher Scientific) from $10 \text{ }^\circ\text{C}$ to $95 \text{ }^\circ\text{C}$ at a rate of $1 \text{ }^\circ\text{C min}^{-1}$. Changes in fluorescence were monitored at excitation and emission wavelengths of $580 \pm 10 \text{ nm}$ and $623 \pm 14 \text{ nm}$, respectively. The melting temperatures (i.e., temperature at which 50% of the protein is unfolded; T_m) of each reaction were determined in the Protein Thermal Shift™ Software v1.4 (Thermo Fisher Scientific). T_m values were calculated for each reaction and compared to the reference T_m values (i.e., solutions consisting of BRPF1b, 1% DMSO, and dye) to obtain ΔT_m values for each compound. To rule out any potential effects caused by the presence of DMSO, an additional control consisting of BRPF1b (no DMSO), and dye was included.

BRPF1b immobilization: Binding affinity (K_D) and kinetics (k_a and k_d) were assessed by GCI on the WAVEdelta system (Creoptix AG, Switzerland). A combined His-tag capture/amine coupling approach was used to generate a stable sensor surface. All flow rates were maintained at $133.3 \mu\text{L/min}$, unless otherwise stated, and dissociation times were maintained at 60 sec throughout. Briefly, a PCH-NTA WAVEchip (Creoptix AG, Switzerland) was conditioned with a 180 sec injection of 0.1 M Borate (pH 9), 1 M NaCl followed by two 180 sec injections of 0.35 M EDTA (pH 8). Excess EDTA was removed following a 60 sec injection with running buffer (10 mM HEPES pH 7.4, 150 mM NaCl and 0.05% (v/v) Surfactant P20) at a flow rate of $33.3 \mu\text{L/min}$. The nitrilotriacetic acid (NTA) on the surface of the WAVEchip was charged with a 420 sec injection of 0.5 M NiCl_2 (Cytiva,

Sweden) and excess NiCl₂ was removed by two 60 sec injections of running buffer. To increase capture stability on the WAVEchip, BRPF1b was amine coupled to the NTA-Ni²⁺ surface. The charged NTA-Ni²⁺ groups were activated with 0.2 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 50 mM N-hydroxysuccinimide (NHS) (Cytiva, Sweden). Next, 1.2 μM BRPF1b, diluted in buffer, was injected for 420 sec and captured via coordination of the His-tag and NTA-Ni²⁺ groups. Any remaining active succinimide esters were quenched by a single 420 sec injection of 1 M ethanolamine hydrochloride-NaOH (pH 8.5) (Cytiva, Sweden) at 15 μL/min. Following another 60 sec wash with running buffer, excess Ni²⁺ ions as well as any remaining BRPF1b that had not been amine coupled were removed with a 420 sec injection of 0.35 M EDTA (pH 8). The reference channel was prepared in the same way, but without injection of BRPF1b. All capture and immobilization steps were carried out at 25°C.

WaveRAPID® GCI screen: The WAVEdelta system allows for kinetic interaction analysis based on a pulsed injection scheme (waveRAPID®) using a single analyte concentration.^[36] This approach allows for fast screening of a large number of samples using six repeated pulse durations ranging from 0.031 sec – 0.5 sec. WaveRAPID® was initially used to confirm that immobilized BRPF1b was able to bind **GSK6853** and **NI-57** and later to screen the virtual hits. All 51 virtual hit compounds were screened using the weak waveRAPID® preset at 100 μM diluted in running buffer (10 mM HEPES pH 7.4, 150 mM NaCl and 0.05% (v/v) Surfactant P20) with a final DMSO concentration of 1% (v/v). Binding was assessed at 15 °C using running buffer supplemented with 1% (v/v) DMSO. WaveRAPID® hits were selected based on the following criteria: association error < 60, dissociation error < 60, R_{max} > 2.0.

MCK GCI screen: For more accurate determination of affinity and kinetics, MCK experiments were performed in duplicate for the 36 compounds selected from waveRAPID® using a 2-fold serial dilution series (100 μM – 1.56 μM). The DMSO concentration of the compound dilutions were maintained at 1% (v/v) and experiments were performed with running buffer supplemented with 1% (v/v) DMSO. WaveRAPID® GCI parameters were maintained for MCK experiments to allow for direct comparison of both sets of results.

Ligand observed NMR: Samples were prepared from a 260 mM protein stock solution in NMR buffer (20 mM NaPi pH 7.5, 150 mM NaCl) and nominally 20 mM compound stock solutions in DMSO-d₆ (Cambridge Isotope Laboratories) to produce final concentrations of 26 μM protein and 200 μM or 400 μM compound diluted in NMR buffer with 5% (v/v) D₂O, 0.00025% (w/v) TSP. To limit protein consumption, some compounds were screened in pairs with design of the mixtures aided by the AnalysisScreen^[58] from the CCPNMR software suite.^[59] Spectra were recorded at 599.73 MHz on a Bruker AVANCE IIIHD spectrometer equipped with a 5 mm TCI cryoprobe.

Supporting information

SI Figure 1: Chemical structure and biological data of the 27 known BRPF1b inhibitors used for the similarity search.

SI Figure 2: Representative waveRAPID® sensorgram.

SI Figure 3: LCMS of Compound 2

SI Figure 4: LCMS of Compound 3

SI Figure 5: LCMS of Compound 4

SI Figure 6: LCMS of Compound 5

SI Figure 7: LCMS of Compound 7

SI Table 1: Comparison of **GSK6853** and **NI-57** binding to BRPF1a and BRPF1b.

Acknowledgements

We would like to thank the Edinburgh Protein Production Facility (EPPF) for producing the BRPF1b protein required to conduct the studies described in this article.

We would like to thank Mrs June Southall at the University of Glasgow Structural Biology and Biophysical Characterization Facility for assistance with preparation of samples for NMR spectroscopy.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Supporting information for this article is given via a link at the end of the document.

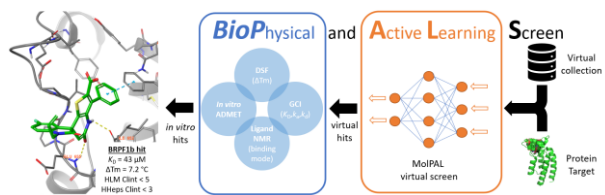
Keywords: Hit identification, Virtual Screening, Machine Learning, BRPF1, Grating Coupled Interferometry

References

- [1] D. G. Brown, J. Boström, *J. Med. Chem.* **2018**, *61*, 9442–9468.
- [2] D. G. Brown, *J. Med. Chem.* **2023**, *66*, 7101–7139.
- [3] H. Berman, K. Henrick, H. Nakamura, *Nat Struct Mol Biol* **2003**, *10*, 980–980.
- [4] S. Subramaniam, *IUCrJ* **2019**, *6*, 1–2.
- [5] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [6] "REAL Database - Enamine," can be found under <https://enamine.net/compound-collections/real-compounds/real-database>, n.d.
- [7] N. Arul Murugan, G. Ruba Priya, G. Narahari Sastry, S. Markidis, *Drug Discovery Today* **2022**, *27*, 1913–1923.
- [8] F. Gentile, J. C. Yaacoub, J. Gleave, M. Fernandez, A.-T. Ton, F. Ban, A. Stern, A. Cherkasov, *Nat Protoc* **2022**, *17*, 672–697.
- [9] F. Potlitz, A. Link, L. Schulig, *Expert Opinion on Drug Discovery* **2023**, *18*, 303–313.
- [10] D. E. Graff, E. I. Shakhnovich, C. W. Coley, *Chem. Sci.* **2021**, *12*, 7866–7881.
- [11] P. Bamborough, H. A. Barnett, I. Becher, M. J. Bird, C. Chung, P. D. Craggs, E. H. Demont, H. Diallo, D. J. Fallon, L. J. Gordon, P. Grandi, C. I. Hobbs, E. Hooper-Greenhill, E. J. Jones, R. P. Law, A. Le Gall, D. Lugo,

- A.-M. Michon, D. J. Mitchell, R. K. Prinjha, R. J. Sheppard, A. J. B. Watson, R. J. Watson, *ACS Med Chem Lett* **2016**, *7*, 552–557.
- [12] N. Igoe, E. D. Bayle, C. Tallant, O. Fedorov, J. C. Meier, P. Savitsky, C. Rogers, Y. Morias, S. Scholze, H. Boyd, D. Cunoosamy, D. M. Andrews, A. Cheasty, P. E. Brennan, S. Müller, S. Knapp, P. V. Fish, *J. Med. Chem.* **2017**, *60*, 6998–7011.
- [13] S. Lapetina, H. Gil-Henn, *J Biol Methods* **2017**, *4*, e62.
- [14] M. Zeilinger, F. Pichler, L. Nics, W. Wadsak, H. Spreitzer, M. Hacker, M. Mitterhauser, *EJNMMI Research* **2017**, *7*, 22.
- [15] N. P. Coussens, T. L. Foley, S. R. J. Hoare, J. McGee, G. S. Sittampalam, in *Assay Guidance Manual [Internet]*, Eli Lilly & Company And The National Center For Advancing Translational Sciences, **2004**.
- [16] M. Glicksman, Z. Li, S. Markossian, T. Riss, G. S. Sittampalam, J. O. Joseph Trask, Y. Wang, in *Assay Guidance Manual [Internet]*, Eli Lilly & Company And The National Center For Advancing Translational Sciences, **2004**.
- [17] P. A. Clemons, *Current Opinion in Chemical Biology* **2004**, *8*, 334–338.
- [18] "Opportunities and challenges in phenotypic drug discovery: an industry perspective | Nature Reviews Drug Discovery," can be found under <https://www.nature.com/articles/nrd.2017.111>, **n.d.**
- [19] "Recent advances in phenotypic drug discovery - PMC," can be found under <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431967/>, **n.d.**
- [20] E. L. Berg, *Cell Chemical Biology* **2021**, *28*, 424–430.
- [21] A. T. Plowright, L. Drowley, in *Annual Reports in Medicinal Chemistry* (Ed.: R. A. Goodnow), Academic Press, **2017**, pp. 263–299.
- [22] T. Kaminski, A. Gunnarsson, S. Geschwindner, *ACS Sens.* **2017**, *2*, 10–15.
- [23] K. Narayan, S. S. Carroll, in *Applied Biophysics for Drug Discovery*, John Wiley & Sons, Ltd, **2017**, pp. 93–105.
- [24] J. M. Rainard, G. C. Pandarakalam, S. P. McElroy, *SLAS Discov* **2018**, *23*, 225–241.
- [25] R. A. Zimmermann, M. Schwickert, J. L. Meidner, Z. Nidoieva, M. Helm, T. Schirmeister, *ACS Pharmacol. Transl. Sci.* **2022**, *5*, 1079–1085.
- [26] D. A. Annis, E. Nickbarg, X. Yang, M. R. Ziebell, C. E. Whitehurst, *Curr Opin Chem Biol* **2007**, *11*, 518–526.
- [27] W. Pitsawong, V. Buosi, R. Otten, R. V. Agafonov, A. Zorba, N. Kern, S. Kutter, G. Kern, R. A. Pádua, X. Meniche, D. Kern, *eLife* **2018**, *7*, e36656.
- [28] L. Galazzo, G. Meier, D. Janulien, K. Parey, D. De Vecchis, B. Striednig, H. Hilbi, L. V. Schäfer, I. Kuprov, A. Moeller, E. Bordignon, M. A. Seeger, *Science Advances* **2022**, *8*, eabn6845.
- [29] J. Frimi, M. Gallei, Z. Gelová, A. Johnson, E. Mazur, A. Monzer, L. Rodriguez, M. Roosjen, I. Verstraeten, B. D. Živanović, M. Zou, L. Fiedler, C. Giannini, P. Grones, M. Hrtyan, W. A. Kaufmann, A. Kuhn, M. Narasimhan, M. Randuch, N. Rydza, K. Takahashi, S. Tan, A. Teplova, T. Kinoshita, D. Weijers, H. Rakusová, *Nature* **2022**, *609*, 575–581.
- [30] C. Hong, N. J. Byrne, B. Zamylny, S. Tummala, L. Xiao, J. M. Shipman, A. T. Partridge, C. Minnick, M. J. Breslin, M. T. Rudd, S. J. Stachel, V. L. Rada, J. C. Kern, K. A. Armacost, S. A. Hollingsworth, J. A. O'Brien, D. L. Hall, T. P. McDonald, C. Strickland, A. Brooun, S. M. Soisson, K. Hollenstein, *Nat Commun* **2021**, *12*, 815.
- [31] F. Andres, L. Iamele, T. Meyer, J. C. Stüber, F. Kast, E. Gherardi, H. H. Niemann, A. Plückthun, *J Mol Biol* **2019**, *431*, 2020–2039.
- [32] H. An, K. T. Elvers, J. A. Gillespie, K. Jones, J. R. Attack, O. Grubisha, T. A. Shelkovnikova, *Nucleic Acids Research* **2022**, *50*, e119.
- [33] U. Hohmann, J. Santiago, J. Nicolet, V. Olsson, F. M. Spiga, L. A. Hothorn, M. A. Butenko, M. Hothorn, *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 3488–3493.
- [34] D. Couto, A. Richter, H. Walter, D. Furkert, M. Hothorn, D. Fiedler, *Biochemistry* **2021**, *60*, 2739–2748.
- [35] E. A. FitzGerald, M. T. Butko, P. Boronat, D. Cederfelt, M. Abramsson, H. Ludviksdottir, J. E. Van Muijlwijk-Koezen, I. J. P. De Esch, D. Dobritzsch, T. Young, U. H. Danielson, *RSC Adv.* **2021**, *11*, 7527–7537.
- [36] Ö. Kartal, F. Andres, M. P. Lai, R. Nehme, K. Cottier, *SLAS Discovery* **2021**, *26*, 995–1003.
- [37] C. L.-H. Cheng, F. H.-C. Tsang, L. Wei, M. Chen, D. W.-C. Chin, J. Shen, C.-T. Law, D. Lee, C. C.-L. Wong, I. O.-L. Ng, C.-M. Wong, *Commun Biol* **2021**, *4*, 888.
- [38] H. Shima, K. Yamagata, Y. Aikawa, M. Shino, H. Koseki, H. Shimada, I. Kitabayashi, *Int J Hematol* **2014**, *99*, 21–31.
- [39] "Acute myeloid leukaemia (AML)," can be found under <https://www.cancerresearchuk.org/about-cancer/acute-myeloid-leukaemia-aml>, **n.d.**
- [40] J. Zhu, A. Caffisch, *J. Med. Chem.* **2016**, *59*, 5555–5561.
- [41] J. Zhu, C. Zhou, A. Caffisch, *Eur J Med Chem* **2018**, *155*, 337–352.
- [42] E. H. Demont, P. Bamborough, C. Chung, P. D. Craggs, D. Fallon, L. J. Gordon, P. Grandi, C. I. Hobbs, J. Hussain, E. J. Jones, A. Le Gall, A.-M. Michon, D. J. Mitchell, R. K. Prinjha, A. D. Roberts, R. J. Sheppard, R. J. Watson, *ACS Med. Chem. Lett.* **2014**, *5*, 1190–1195.
- [43] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Advanced Drug Delivery Reviews* **1997**, *23*, 3–25.
- [44] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, *J. Am. Chem. Soc.* **1997**, *119*, 12257–12261.
- [45] M. Mayer, B. Meyer, *Angewandte Chemie International Edition* **1999**, *38*, 1784–1788.
- [46] "WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability | SpringerLink," can be found under <https://link.springer.com/article/10.1023/A:1013302231549>, **n.d.**
- [47] F. H. Larsen, H. J. Jakobsen, P. D. Ellis, N. Chr. Nielsen, *J. Phys. Chem. A* **1997**, *101*, 8597–8606.
- [48] "ChEMBL Database," can be found under <https://www.ebi.ac.uk/chembl/>, **n.d.**
- [49] eMolecules, "eMolecules," can be found under <https://www.emolecules.com>, **n.d.**
- [50] D. Bajusz, A. Rácz, K. Héberger, *Journal of Cheminformatics* **2015**, *7*, 20.
- [51] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J Comput Aided Mol Des* **2013**, *27*, 221–234.
- [52] O. Trott, A. J. Olson, *Journal of Computational Chemistry* **2010**, *31*, 455–461.
- [53] "AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings | Journal of Chemical Information and Modeling," can be found under <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00203>, **n.d.**
- [54] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, *Nat Chem* **2012**, *4*, 90–98.
- [55] "RDKit," can be found under <https://rdkit.org/>, **n.d.**
- [56] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107–113.
- [57] M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, F. R. Salemme, *J Biomol Screen* **2001**, *6*, 429–440.
- [58] L. G. Mureddu, T. J. Ragan, E. J. Brooksbank, G. W. Vuister, *J Biomol NMR* **2020**, *74*, 565–577.
- [59] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, *J Biomol NMR* **2016**, *66*, 111–124.

Entry for the Table of Contents



Traditional hit identification requires tailoring to each biological target and is reliant on multiple target-specific assay technologies. Combining Molecular Pool-based Active Learning (MoIPAL) with a suite of *in vitro* biophysical methods, **BioIPAL** shifts this paradigm by providing a standardized and data-rich hit identification platform applicable to most biological targets. The application to BRPF1b afforded a range of high-quality starting points.