



**HAL**  
open science

# **Integrated Strategy for Lead Optimization Based on Fragment Growing: The Diversity-Oriented-Target-Focused-Synthesis Approach**

Laurent Hoffer, Yuliia Voitovich, Brigitt Raux, Kendall Carrasco, Christophe Muller, Aleksey Fedorov, Carine Derviaux, Agnès Amouric, Stéphane Betzi, Dragos Horvath, et al.

## ► To cite this version:

Laurent Hoffer, Yuliia Voitovich, Brigitt Raux, Kendall Carrasco, Christophe Muller, et al.. Integrated Strategy for Lead Optimization Based on Fragment Growing: The Diversity-Oriented-Target-Focused-Synthesis Approach. *Journal of Medicinal Chemistry*, 2018, 61 (13), pp.5719-5732. <10.1021/acs.jmedchem.8b00653>. <hal-02067797>

**HAL Id: hal-02067797**

**<https://hal.science/hal-02067797v1>**

Submitted on 6 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# An Integrated Strategy for Lead Optimization based on Fragment Growing: The DOTS (Diversity-Oriented Target-focused Synthesis) Approach

*Laurent Hoffer<sup>a,1</sup>, Yuliia V. Voitovich<sup>a,b,1</sup>, Brigitt Raux<sup>a</sup>, Kendall Carrasco<sup>a</sup>, Christophe Muller<sup>d</sup>, Aleksey Y. Fedorov<sup>b</sup>, Carine Derviaux<sup>d</sup>, Agnès Amouric<sup>a,d</sup>, Stéphane Betzi<sup>a</sup>, Dragos Horvath<sup>c</sup>, Alexandre Varnek<sup>c</sup>, Yves Collette<sup>a,d</sup>, Sebastien Combes<sup>a</sup>, Philippe Roche<sup>a,\*</sup> and Xavier Morelli<sup>a,d,\*</sup>*

<sup>a</sup>CRCM, CNRS, INSERM, Institut Paoli-Calmettes, Aix-Marseille Univ, Marseille, France

<sup>b</sup>Department of Organic Chemistry, Lobachevsky State University of Nizhni Novgorod, 23 Gagarin Avenue, Nizhni Novgorod, 603950, Russia

<sup>c</sup>Laboratoire de Chemoinformatique, CNRS UMR7140, 1 rue Blaise Pascal, 67000, Strasbourg, France

<sup>d</sup>IPC Drug Discovery Platform, Institut Paoli-Calmettes, 232 Boulevard de Sainte-Marguerite, 13009, Marseille, France

<sup>1</sup>Laurent Hoffer and Yuliia Voitovich contributed equally to this work

\*To whom correspondence should be addressed. E-mail: [philippe.roche@inserm.fr](mailto:philippe.roche@inserm.fr)  
[xavier.morelli@inserm.fr](mailto:xavier.morelli@inserm.fr)

**Abstract**

Over the past few decades, hit identification has been greatly facilitated by advances in high-throughput and fragment-based screenings. One major hurdle remaining in drug discovery is process automation of hit-to-lead (H2L) optimization. Here, we report a time and cost-efficient integrated strategy for H2L optimization and partially automated design of potent chemical probes consisting of focused chemical library design and virtual screening coupled with robotic diversity-oriented *de novo* synthesis and automated *in vitro* evaluation. The virtual library is generated by combining an activated fragment corresponding to the substructure binding to the target with a collection of functionalized building blocks using *in silico*-encoded chemical reactions carefully chosen from a list of one-step organic transformations that are relevant in medicinal chemistry. The proof of concept was demonstrated using the optimization of bromodomain inhibitors as a test case, leading to the validation of several compounds with affinity improved by several orders of magnitude.

**Keywords:**

Hit-to-lead, fragment, optimization, growing, focused library, diversity-oriented library, virtual screening, *de novo* synthesis, drug discovery, integrated pipeline, automation.

## Introduction

The initial stage of the drug discovery process, referred to as “hit identification”, involves identifying fragment(s) or molecule(s) able to bind to a biological target in the micro- to millimolar affinity range.<sup>1, 2</sup> The development of high-throughput screening (HTS) approaches, particularly automation and miniaturization processes, and progress in chemical library design have greatly improved hit identification by allowing the screening of large high-quality collections of compounds.<sup>3</sup> In addition, fragment-based approaches have emerged in the recent years as an alternative strategy to identify weakly interacting fragments from small diverse libraries of compounds.<sup>4, 5</sup>

Hits found by HTS or fragment-based approaches are usually evaluated and validated through a series of *in vitro* and/or *in cellulo* assays. These low affinity compounds then need to be optimized into probes and/or drugs using hit- to-lead (H2L) approaches to develop more potent and selective lead compounds that can enter preclinical evaluation. Several hits are usually identified for each target, making it crucial to develop cost-effective strategies to rapidly propose optimized compounds from a given hit. However, improving hit compounds to potent and selective molecular probes or drug candidates can be hazardous, time consuming and expensive. Thus, overcoming these bottlenecks in hit optimization and developing strategies to accelerate, improve and automate the process from hit identification to lead or probe optimization are of primary importance.<sup>6, 7</sup>

This optimization phase to improve the affinity, selectivity and physicochemical properties of the compounds involves a combination of experimental and *in silico* approaches.<sup>8, 9</sup> Structure-activity relationship (SAR) by catalog is the first straightforward step that can be employed to explore the chemical space around the hit using similar commercially available compounds. This approach is limited by the availability of the compounds and by their diversity. The next step is chemical SAR studies, in which chemists systematically explore all positions of the compounds suitable for modifications. Synthesized compounds are experimentally evaluated to drive the sequential synthetic steps. Several cycles are usually needed to reach the affinity values required for probes or drug-like candidates. Ligand-based *in silico* approaches have been developed to assist this process

1  
2  
3 and improve hit compounds, including 2D or 3D quantitative structure-activity relationships  
4 (QSAR).<sup>10, 11</sup> Finally, structure-based approaches such as molecular docking and virtual screening (VS)  
5 can significantly accelerate compound development by considering structural information about the  
6 target.<sup>12</sup> This structural information greatly facilitates the H2L optimization by analyzing the binding  
7 mode of the hit and identifying nearby protein subpockets that could be filled. For instance, the  
8 strategy generally implemented is known as the “growing phase”, which involves the design of  
9 putative optimized compounds that could fit the subpocket while conserving the original interactions  
10 of the fragment within the binding site.<sup>13, 14</sup>

11  
12  
13  
14  
15  
16  
17  
18  
19  
20 As mentioned previously, SAR by catalog is limited to commercially available compounds. For the  
21 fragment growing strategy, it is generally necessary to be able to synthesize small-size chemical  
22 libraries focused around the initial fragment. Several approaches, involving coupling an activated  
23 form of the fragment with a series of compatible building blocks (BBs), have been developed to  
24 design such virtual libraries.<sup>15, 16</sup> However, it is crucial to improve the quality of these chemical  
25 libraries by taking into account the physicochemical properties of the compounds and the simplicity  
26 and efficiency of their synthesis. The compounds must possess medicinal chemistry and ADME  
27 properties that are compatible with their development as probes or drugs.

28  
29  
30  
31  
32  
33  
34  
35  
36  
37 Synthetic approaches, such as diversity-oriented synthesis (DOS), have been developed to generate  
38 structural diversity in a rational and efficient manner. The goal of such strategies is to maximize the  
39 chemical space explored while limiting the number of compounds to a minimum.<sup>17, 18</sup> DOS concept  
40 has been applied with success for the discovery of biologically active compounds and their evolution  
41 as probes.<sup>19-25</sup>

42  
43  
44  
45  
46  
47  
48 In this study, we report a H2L optimization strategy integrating computational and automated  
49 experimental methods: DOTS (Diversity-Oriented Target-focused Synthesis). A general pipeline has  
50 been implemented to design more potent chemical probes starting from an initial fragment. First,  
51 the binding mode of the initial hit is identified using X-ray crystallography. Then, the *in silico*  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 optimization strategy relies on 2 main steps: i/ the design of a diversity-oriented chemical library  
4 around the initial hit using medicinal chemistry-relevant reactions and a collection of commercially  
5 available BBs and ii/ the VS of this chemical library using the conformational sampling tool S4MPLE to  
6  
7 select the best interacting compounds.<sup>26</sup> These compounds are prioritized using interaction energy  
8  
9 ranking, followed by synthesis and evaluation using robotic platforms. The experimental process  
10  
11 (synthesis and evaluation) can be set up and validated using the same plates without any  
12  
13 intermediary purification processes. The entire cycle of optimization can be repeated iteratively to  
14  
15 design more potent and selective compounds (Fig. 1).  
16  
17  
18

19  
20 DOTS was successfully applied to optimize a recently characterized inhibitor of bromodomain (BRD),  
21  
22 a protein from the bromodomain and extra-terminal (BET) family.<sup>27</sup> Several new compounds  
23  
24 exhibiting improved affinities ( $IC_{50}$  and/or  $K_D$  values) have been synthesized and validated starting  
25  
26 from the core fragment. The X-ray structure of the most potent compound was solved and used to  
27  
28 decipher the molecular basis of the improved affinity.  
29  
30

## 31 **Results and Discussion**

32  
33 The DOTS strategy involves several integrated steps summarized in Fig. 1. It combines highly  
34  
35 automated computational and experimental approaches for rapid and cost-effective optimization of  
36  
37 hit compounds. The *in silico* step includes virtual chemical library design and structure-based VS (Fig.  
38  
39 2). The experimental methods involve diversity-oriented *de novo* synthesis and *in vitro* evaluation  
40  
41 with robotic platforms.  
42  
43

44  
45 The minimum common substructure necessary to bind the target is identified using either fragment-  
46  
47 based approaches or deconstruction of an identified hit compound.<sup>5, 28</sup> The binding mode is  
48  
49 characterized by structural models of the fragment complexed with the biological target using X-ray  
50  
51 crystallography (Fig. 1.1). Other biophysical methods could also be used at this stage, such as NMR or  
52  
53 cryo-EM.<sup>29, 30</sup> The common core is used to build a virtual chemical library focused on the hit  
54  
55 compound (Figs. 1.2 and 1.3). The design of this diversity-oriented chemical library is achieved by  
56  
57  
58  
59  
60

1  
2  
3 coupling an activated form of the core fragment with a diverse collection of commercially available  
4 functionalized BBs. Prioritized compounds are selected through VS (Figs. 1.4 and 1.5). The  
5 corresponding BBs are then purchased (Fig. 1.6), followed by parallel synthesis (Fig. 1.7) and *in vitro*  
6 evaluation of the compounds (Fig. 1.8).  
7  
8  
9

### 10 11 **Building virtual focused libraries**

12 Progresses in computational methods have made it possible to screen millions of virtual compounds  
13 *in silico* in a reasonable amount of time. However, computer-generated compounds are not always  
14 commercially available or easily accessible to organic synthesis. It is therefore crucial to incorporate  
15 well-recognized medicinal chemistry rules in the design of virtual libraries.<sup>31, 32</sup>  
16  
17  
18  
19

20  
21  
22 DOTS uses a relevant set of robust organic synthesis reactions defined by Hartenfeller and coworkers  
23 or implemented in house.<sup>16</sup> These rules, designed by both medicinal chemists and molecular  
24 modelers, include the most common and well-accepted reactions encountered during the H2L stages  
25 in the pharma industry. Most of these rules involve only a one-step reaction, and half of them  
26 concern cycle formation. The first step of virtual library design is the careful choice of one or a few  
27 chemical reactions compatible with the starting material from the list of available chemical reactions  
28 (Figs. 1.2 and 2.A). Consequently, the initial hit must be activated to include a reactive functionality  
29 that matches at least one reaction in the set of medicinal chemistry-relevant reactions. A collection  
30 of functionalized BBs compatible with the selected reaction(s) is assembled from commercially  
31 available databases. The activated hit is coupled *in silico* to the different BBs to generate the focused  
32 chemical library (Figs. 1.3 and 2.A). This strategy almost guarantees that compounds generated in the  
33 virtual library are easily amenable to organic synthesis in one or two steps with high expected yields  
34 and no byproducts. In addition, all compounds in the virtual library contain the original hit scaffold as  
35 a submoiety, making it a focused library. Several fully automated post-processing stages are applied  
36 to the virtual library to extract a diverse subset of duplicate-free representative compounds with a  
37 focus on medicinal chemistry-relevant structures with reasonable physiochemical properties (Fig.  
38 2.B, see supporting information).  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Conformational sampling with S4MPLE

The subsequent VS step involves constrained sampling with S4MPLE (Fig. 2.C), a molecular modeling tool that relies on a Lamarckian genetic algorithm for conformational sampling and force-field formalism for energy assessment.<sup>33</sup> S4MPLE uses the AMBER energy force field<sup>34</sup> and GAFF<sup>35</sup> to simulate proteins and small organic molecules, respectively.

The VS stage is performed under restraints to maintain the binding mode of the original hit according to the generic hit growing paradigm.<sup>13</sup> The rest of the molecule can move freely to explore the surrounding pockets. Following the sampling process, the energy of all nonredundant poses is minimized without applying any restraints. Compounds displaying a shift larger than a user-defined root-mean-square deviation (RMSD) threshold after energy minimization are discarded. The relative binding energy of each ligand is estimated by computing the energy difference of the best pose in the bound and free states. Compounds are then ranked according to this computed energy. The steps leading to the ranking of the compounds are completely automated. The final selection of compounds is achieved after visual inspection of the best poses; however, this last step is limited to a small number of poses, leading to a significant reduction in time.

### Automated synthesis platform

To enable faster and more efficient preparation of an extensive series of chemical derivatives, an automated high-throughput robotic platform can be used.<sup>36,37</sup> These platforms are powerful systems that allow the rational design of a diverse library of small molecules by performing up to 96 reactions in parallel. It dramatically accelerates the synthesis of a broad spectrum of novel derivatives and the optimization of the physical and biological properties of molecules for rapid generation of SAR. This automated strategy does not require significant changes in the synthetic protocols available, which allows fast adaptation of "hands-on" synthetic procedures to automated programs. Multiwell synthesizer robots are useful for synthesizing a range of compounds via a single type of one-step reaction. This strategy is perfectly suited to prepare the compounds originating from the virtual library and selected by the *in silico* VS approach. Notably, in this approach, only those compounds

1  
2  
3 that are predicted to bind to the biological target are synthesized as opposed to standard  
4  
5 combinatorial chemistry approaches.

### 6 7 **Optimization of BRD inhibitors with DOTS: proof of concept**

8  
9 BRDs are protein domains that specifically recognize acetylated lysine (Kac) residues, such as those  
10  
11 present on the histone tails. The BRD4 protein contains two BRD in tandem and belongs to the  
12  
13 bromodomain and extra-terminal (BET) family.<sup>38</sup> These epigenetic readers of lysine acetylation  
14  
15 emerged as druggable targets due to their role in inflammatory and cardiovascular diseases and  
16  
17 cancer.<sup>39</sup> Thus, developing molecules that could efficiently disrupt this protein-protein interaction is  
18  
19 highly desirable.

20  
21  
22 We have recently reported xanthine-containing derivatives identified through medium-throughput  
23  
24 screening as selective inhibitors of the first BRD of the human BRD4 protein, namely, BRD4(BD1).<sup>27</sup>  
25  
26 The binding mode was solved using X-ray crystallography. SAR studies revealed the minimum scaffold  
27  
28 necessary for this binding, and analysis of 3D structures obtained in complex with different  
29  
30 derivatives revealed surrounding pockets that could be explored by VS.

31  
32  
33 **Choice of chemical reaction:** Based on already synthesized intermediates, a sulfonamidation reaction  
34  
35 was selected for the design of the virtual chemical library (Fig. S1). For this condensation, both amino  
36  
37 and chlorosulfonyl reactive groups divided between the substrate and reagent are required. We  
38  
39 solved the X-ray structure of the “common-core” fragment **F1** in complex with the target (PDB code:  
40  
41 6FNX), which revealed a nearby subpocket delimited by 5 residues (F79, D145, L148, M149 and  
42  
43 A152), most of which have a hydrophobic side chain (see Fig. S2 for the 2D structure of **F1**). Modeling  
44  
45 studies revealed that substitutions in the para position of the benzyl group should provide better  
46  
47 access to the subpocket.

48  
49  
50 **Design of chemical library:** A primary methyleneamine was introduced to the initial fragment **F1** as a  
51  
52 nucleophilic functionality. The methylene group was introduced as a spacer to integrate a bend into  
53  
54 the structures to satisfy the steric constraints of the binding site. This benzylamine-containing  
55  
56

1  
2  
3 compound **F2** served as an activated starting material for the *de novo* design (Figs. S1 and S2). **F2** was  
4  
5 virtually combined with sulfonyl chloride-containing BBs to generate the virtual library built around  
6  
7 fragment **F1**. This library was post-processed and ultimately contained 576 compounds (see  
8  
9 supporting information for more details).

10  
11 **VS with S4MPLE:** The final subset of 576 compounds was virtually screened using S4MPLE with  
12  
13 constraints on the sole xanthine moiety. This substructure interacts with conserved N140, a known  
14  
15 hot spot of BRD4(BD1); this fundamental interaction was preserved during the H2L process. The  
16  
17 remaining parts of the molecules, namely from the original phenyl group to the newly added atoms,  
18  
19 were allowed to explore adjacent subpockets on the protein surface. At the end of the process, the  
20  
21 energy of the compounds was minimized without any restraints. Importantly, the VS routine is  
22  
23 embedded with the chemical library design, and the entire process is fully automated.

24  
25  
26 **Selection of compounds to be synthesized:** The prioritization of the putative optimized compounds  
27  
28 was based on several criteria, the most important of which was the force-field energy difference  
29  
30 between the bound and free states of the compound. Only the top 100 compounds were considered  
31  
32 for further analysis. Among those, compounds that did not converge towards the targeted pocket  
33  
34 area (F79, D145, L148, M149 and A152) and compounds exhibiting a shift from the initial binding  
35  
36 mode were not considered. We performed a clustering of compounds based on structural similarities  
37  
38 and selected only one representative compound for each cluster (see supplementary information).  
39  
40 Remaining poses were subjected to visual inspection. When no additional strong interaction was  
41  
42 detected, the less substituted compound was selected for the synthesis stage. At the end of this  
43  
44 process, 17 representative molecules were selected for the subsequent synthesis stage (Fig. S2 and  
45  
46 Table 1). Each of these compounds did not necessarily correspond to the best-ranked one within a  
47  
48 given cluster but was predicted to have the highest ligand efficiency value.<sup>40</sup>

49  
50  
51  
52 **Robotic synthesis:** The 17 compounds were generated using a sulfonylation reaction of benzylamino-  
53  
54 containing xanthine derivative **F2** with appropriate sulfonyl chloride BBs (Fig. S1). All compounds  
55  
56  
57  
58  
59  
60

1  
2  
3 have been efficiently prepared in high yields ( $87 \pm 11\%$ ) using our automated robotic platform from  
4  
5 Chemspeed.

### 6 7 ***In vitro* evaluation of the compounds**

8  
9 The crude synthetic products were directly transferred to a Labcyte Access/Echo<sup>®</sup> Laboratory  
10  
11 Workstation to assess the compounds for their ability to disrupt BRD/histone complexes using a  
12  
13 homogeneous time-resolved fluorescence (HTRF)-based assay.<sup>41</sup> The primary screening results  
14  
15 revealed that, compared with the original hit, all 17 *de novo*-designed molecules displayed improved  
16  
17 pIC<sub>50</sub> (Table 1). The initial hit (**F1**) exhibited a pIC<sub>50</sub> value of 4.9, whereas the amino activated  
18  
19 fragment (**F2**) demonstrated a pIC<sub>50</sub> value of 4.6. Fourteen compounds (82%) exhibited an affinity  
20  
21 improvement greater than 1 log, which is the expected range of improvement in a classical round of  
22  
23 hit optimization. They all contained an aromatic ring directly attached to the sulfonyl group. Among  
24  
25 the most improved compounds in terms of affinity were biphenyl derivative (**1**) and 4-  
26  
27 benzyloxyphenyl-sulfonamide derivative (**7**) with pIC<sub>50</sub> values 6.4 and 6.5, respectively.  
28  
29 Tetrahydroquinolinone derivative **9** exhibited high inhibitory activity (pIC<sub>50</sub> = 6.4) whereas its  
30  
31 benzazepin-2-one analog (**17**) exhibited the best target affinity with a pIC<sub>50</sub> value of 6.6, an  
32  
33 approximately 2-log improvement over the initial hit **F1**.  
34  
35

36  
37 In summary, several high affinity inhibitors were characterized starting from a moderate affinity hit  
38  
39 using a fully integrated pipeline combining molecular modeling, chemistry and biochemistry while  
40  
41 relying on a robot-based technology. All selected compounds showed better affinity values than the  
42  
43 initial fragment with an average improvement of 1.2 logs.  
44

### 45 46 **Orthogonal validation of the compounds**

47  
48 To further validate the approach, all compounds exhibiting submicromolar affinities in HTRF were  
49  
50 resynthesized using a state-of-the-art approach, including a purification step, and analyzed using ITC  
51  
52 and TSA as orthogonal validation methods (Table 2). All thirteen purified compounds were confirmed  
53  
54 as high affinity binders by both methods (Fig. S3). Submicromolar K<sub>D</sub> values consistent with the HTRF  
55  
56 data were measured for all but one compound, and the best compound (**17**) showed an affinity of  
57  
58  
59  
60

1  
2  
3 190 nM for its target, corresponding to an affinity improvement that was 60-fold greater than that of  
4  
5 the initial fragment **F1** (Fig. 3).

6  
7 Beside optimization of the ligand affinity for the target, it is important to take into account ADME  
8  
9 parameters during the H2L process. During the preparation of the virtual library several filters were  
10  
11 applied using molecular descriptors that are important to predict the drug-likeness of compounds.  
12  
13 This increases the probability that the designed compounds possess physicochemical properties that  
14  
15 are compatible with further developments. Besides, different efficiency metrics have been proposed  
16  
17 to estimate the potency of compounds during the H2L process.<sup>42</sup> Among them, the ligand efficiency  
18  
19 (LE), defined as the binding free energy per heavy atom count ( $LE = \Delta G/HAC$ ). All 13 validated  
20  
21 compounds show decreased LE values compared to the initial fragment (Table S1). It is not  
22  
23 uncommon in fragment to lead optimization programs to observe such drops in LE values.<sup>43, 44</sup> The  
24  
25 best LE (0.27 kcal/mole/heavy atom) corresponds to 4-methyl-benzene sulfonamide derivative (**13**),  
26  
27 the less substituted compound. However, all compounds exhibit similar LE values ( $0.24 \pm 0.01$   
28  
29 kcal/mole/heavy atom) and therefore this metrics cannot be used to prioritize compounds. The  
30  
31 lipophilic ligand efficiency (LLE) is an estimate of the contribution of lipophilicity to potency ( $LLE =$   
32  
33  $pIC_{50} - cLogP$ ). LLE has been shown to improve during fragment to lead optimization programs and  
34  
35 suggested as a valid metrics to prioritize compounds.<sup>42-47</sup> More than half of the validated compounds  
36  
37 demonstrate better LLE compared to the initial hit **F1** (Table S1) with compounds **8**, **9** and **17** being  
38  
39 the most improved. Finally, the aqueous solubility of the validated compounds has been measured  
40  
41 (Table S1). The most soluble compounds are **17**, **8**, **9**, **13** and **3**. Taking into account these different  
42  
43 results, **17** appears the most appealing compound for further developments although several other  
44  
45 compounds show interesting potential.

46  
47  
48  
49  
50 The X-ray structure of the most improved compound in terms of binding affinity (**17**) complexed with  
51  
52 the BRD4(BD1) protein target was solved (PDB code: 6FO5) to confirm the binding mode predicted by  
53  
54 the modeling experiments and to identify the new interactions responsible for the improvement in  
55  
56 affinity. The final  $2Fo - Fc$  omit map showed clear and unambiguous density throughout the ligand  
57  
58  
59  
60

1  
2  
3 (Fig. S4). As expected, analysis of the X-ray structure revealed that the binding mode of the *N*-  
4 ethylxanthine moiety was conserved (Fig. S5). Comparison of the X-ray structures of **F1** and the most  
5 improved compound (**17**) allowed us to decipher the molecular basis of the improved affinity (Fig. 4).  
6  
7 Direct and water-mediated hydrogen bond interactions are observed with N140 and Y97,  
8  
9 respectively. The network of 5 structural water molecules stabilizing the protein-ligand interaction  
10  
11 and located deep in the Kac binding pocket is also preserved. A slight translation of the xanthine core  
12  
13 was visible between fragment **F1** and compound **17** because the latter molecule cannot penetrate as  
14  
15 deeply into the Kac cavity. As a consequence, structural water molecules are also slightly displaced.  
16  
17  
18 The flexible ZA loop, which plays a key role in target affinity and specificity, is displaced in the  
19  
20 structure of BRD4(BD1) in the presence of compound **17**, resulting in the shift of P95 and D96 by 3.9  
21  
22 Å and 4.9 Å, respectively (Fig. S6). Several new interactions are detected with compound **17** that  
23  
24 could account for the affinity that was better than that of **F1**. Water-mediated hydrogen bond  
25  
26 interactions involving W81 and D145 are detected between BRD4(BD1) and compound **17**. This  
27  
28 compound also exhibits additional Van der Waals contacts with F79, D145 and M149 that are absent  
29  
30 in the initial fragment.  
31  
32  
33

## 34 35 **Conclusion**

36  
37 We report an original integrated strategy that is useful for exploring the chemical space around a  
38  
39 starting hit to optimize its affinity in a time-efficient and cost-effective manner. This approach uses  
40  
41 focused chemical library design and VS methodologies in combination with experimental methods,  
42  
43 including compound synthesis, *in vitro* evaluation and validation.  
44  
45

46 The general DOTS strategy can be summarized as follows (Fig. 1): i) hit identification and  
47  
48 characterization of binding mode, ii) design of a virtual library using medicinal chemistry rules, iii)  
49  
50 selection of the best compounds using structure-based VS, iv) purchase of BBs, v) robotic parallel  
51  
52 synthesis of selected compounds, and vi) *in vitro* evaluation. Most steps are now fully automated.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 The *in silico* stage relies on chemical knowledge and allows the production of accessible and diverse  
4 compounds while exploring the chemical space around the reference hit and matching regular  
5 physicochemical and medicinal chemistry-like features. The properties of the chemical library can be  
6 tuned using different cut-off values for standard molecular descriptors. The next VS phase identifies  
7 putative compounds that fit new pockets in the binding site while maintaining the original  
8 interactions.

9  
10  
11  
12  
13  
14  
15  
16 Following these *in silico* phases, wet-lab stages rely on a robotic platform to synthesize and screen  
17 the compounds predicted to bind to the protein target. The full process can be set up without any  
18 intermediary purification processes, allowing the same 96-well plates to be used in both stages for  
19 maximal efficiency. In practice, in a round of optimization, compounds would be first evaluated  
20 without undergoing any purification process, and only the most active compounds would then be  
21 purified for validation and characterization.

22  
23  
24  
25  
26  
27  
28  
29 As a proof of concept, DOTS has been validated using the optimization of a BRD inhibitor as a test  
30 case, leading to the development of several optimized compounds starting from a unique hit  
31 fragment. All selected compounds exhibited better affinity than the original hit and several displayed  
32 properties compatible with further development. The best optimized compound demonstrated a  
33 greater than 60-fold improvement in its affinity for the BRD4(BD1) target.

34  
35  
36  
37  
38  
39  
40 The DOTS strategy offers several advantages, including speed, effectiveness and novelty. Automation  
41 and standardization of the different steps are intended to improve and accelerate the process rather  
42 than to increase the number of compounds screened, synthesized and evaluated. The shorter  
43 reaction and purification times minimize transfer loss by using the same plates. Standardization of  
44 protocols is also important in making the process more reliable and the results more reproducible.

45  
46  
47  
48  
49  
50  
51 The synergistic approach of DOTS requires an optimal combination of expertise from modelers,  
52 chemists and biologists to better explore the chemical space around given hits. Human medicinal and  
53 organic chemistry expertise is needed to select relevant chemical reactions, whereas computational  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 methods make it possible to limit the number of compounds to be synthesized and evaluated only to  
4  
5 those that are predicted to bind to the target efficiently.  
6

7  
8 The DOTS strategy is perfectly suited for the efficient design of probes with properties compatible  
9  
10 with medicinal chemistry developments. This approach is applicable to targets compatible with *in*  
11  
12 *vitro* automated assays and it is currently applied to a variety of targets which will allow to better  
13  
14 define its applicability domain. Other improvements including the design of covalent or more  
15  
16 complicated compounds are under development.  
17

## 18 19 **Experimental Section.**

### 20 21 **Preparation of building blocks**

22  
23 Preparation of the BB database relies on an automated pipeline illustrated by Fig. S7. First, BBs were  
24  
25 retrieved from the MolPort database (<http://www.molport.com>). This database is updated monthly  
26  
27 and can be easily interrogated using queries related to properties such as structure, price, supplier or  
28  
29 availability using a Java-based API. The retrieved BBs were prepared using a set of tools from  
30  
31 ChemAxon (<http://www.chemaxon.com>). BBs were subjected to common standardization including  
32  
33 salts removal, neutralization and Kekule representation for aromatic systems using Standardizer  
34  
35 (*version 15.5.4.0*). A check of the structures was then performed with StructureChecker to eliminate  
36  
37 compounds with warnings. BBs that contain undesired elements (other than C, H, N, O, P, S, F, Cl, Br,  
38  
39 I, B or Sn) or more than one undefined stereocenter were rejected at this stage. BBs that were  
40  
41 considered too expensive (e.g., more than \$100 for 50 mg) were also discarded. Additional filters  
42  
43 were applied using common physico-chemical properties (heavy atom count  $\leq 24$ , rotational bonds  
44  
45  $\leq 16$ , ring size  $\leq 7$ , ring count  $\leq 3$ , ring system size  $< 3$ ). These values were selected to discard drug-like  
46  
47 compounds from putative BBs while keeping stannane and boronic acid-containing BBs that may  
48  
49 include flexible hydrophobic tails. BBs containing uncommon or exotic substructures that are not  
50  
51 expected in a hit-2-lead context (e.g., adamantane, spiro-ring, etc.) were removed using jsearch  
52  
53 from Jchem Base and a set of SMARTS patterns. Relevant tautomers of the BBs were then computed  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 using cxcalc calculator, a plugin from ChemAxon. Tautomers were computed at this stage and not on  
4  
5 the final focused library to preserve the predefined tautomer of the hit compound in the binding site.  
6  
7 Finally, duplicates were removed during the last stage to generate a collection of 93k preprocessed  
8  
9 BBs.

### 11 **Chemical reactions encoding**

12  
13  
14 The 58 chemical reactions described by Hartenfeller encoded in SMARTS format were retrieved.<sup>16</sup>  
15  
16 Most of these rules are designed as “Reactant1 + Reactant2 → Product” and half of them correspond  
17  
18 to cycle formation. For the bromodomain test case, reaction corresponding to the coupling of a  
19  
20 sulfonylchloride and free amino groups to form a sulfonamide was selected (Fig. S1.A). In house  
21  
22 reactions have also been implemented to handle alternative synthesis schemes in order to design  
23  
24 putative suicide inhibitors involving a cysteine residue.  
25

### 27 **Focused chemical library design**

28  
29 An in house python script, based on the RDKit toolkit (<http://www.rdkit.org>) was used to create the  
30  
31 raw focused library in SMILES format. BBs containing a sulfonyl-chloride group were automatically  
32  
33 retrieved from the pre-processed MolPort database. Compound **F2** containing the primary amine  
34  
35 reactive group, was coupled to the ensemble of 1200 sulfonylchloride-containing BBs to generate the  
36  
37 virtual library (Fig. S1.B). The post-processing of this raw focused library is described in the Fig. 2.B. A  
38  
39 structural check of generated compounds was performed with StructureChecker to eliminate  
40  
41 compounds with warnings. Putative duplicates and compounds that contain at least one undesired  
42  
43 element (other than C, H, N, O, P, S, F, Cl, Br or I) were also rejected. The next filtering step is based  
44  
45 on physico-chemical properties in order to discard structures with obvious warnings. Common  
46  
47 threshold values were used ( $cLogP \leq 5$ ,  $tPSA \leq 150 \text{ \AA}^2$ , rotational bonds  $\leq 10$  and ring system size  $\leq 2$ ).  
48  
49 Due to both the size of original **F2** fragment and the limited number of sulfonylchloride-containing  
50  
51 BBs, a greater value for the maximal molecular weight was employed (575 Da by contrast to the  
52  
53 default 500 Da value). Undesired compounds were excluded based on yuck filters used to generate  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the clean subset of the ZINC database (<http://zinc.docking.org>). Putative PAINS were identified and  
4 removed from the library using rules extracted from FafDrugs.<sup>48</sup> Due to the limited size of the current  
5 chemical library, the search for privileged scaffolds was not applied and a soft clustering approach  
6 was performed to generate the final focused library. In practice, only bromine and iodine containing  
7 analogs of chloride containing compounds were discarded. The final focused library contained 576  
8 compounds.

9  
10  
11 Prior to the VS stage, major microspecies, atomic types, partial charges and one 3D-conformer of  
12 each compound in the virtual library were precomputed. Atoms from the reference structure N-ethyl  
13 xanthine moiety, constrained during subsequent VS, were flagged using a MCS algorithm. Finally, the  
14 generated conformers were automatically superimposed to the reference substructure, which was  
15 prepositioned in the binding site. These steps were performed using in house tools relying on the  
16 ChemAxon Java API (<http://www.chemaxon.com>). GAFF force-field atomic types were mapped using  
17 programs from AmberTools (<http://ambermd.org>).

### 31 **Conformational sampling**

32  
33 Conformational search with S4MPLE was performed as previously described.<sup>26, 33, 49</sup> Initial X-ray  
34 structure of the bromodomain target (PDB code 5EGU) was prepared using MOE version 2016  
35 (Chemical Computing Group Inc., Montreal, QC, Canada). Protonate3D tool was used to add  
36 hydrogen. A probe atom was positioned in the target binding pocket (F79, D145, L148, M149 and  
37 A152). All residues, with at least one atom within 10 Å radius from either N-ethyl xanthine core or  
38 probe atom, were selected to define the binding site. A large binding site was defined because  
39 S4MPLE relies on a FF-based energy function. The sampling stage mainly consisted in 3 independent  
40 simulations of 400 generations with a population of 30 individuals. Atoms from the N-ethyl-xanthine  
41 core were constrained during this stage, while all other ligand atoms were allowed to move freely. All  
42 saved poses were merged into one file before switching to the post-processing stage that involved a  
43 minimization of all non-redundant poses while unlocking all ligand atoms (by removing all  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 constraints). This minimization step also included systematic optimization of polar hydrogen from  
4 hydroxyl groups of both ligand and binding site. The energy used for ranking was defined as the  
5 lowest potential energy of the complex minus the lowest potential energy of the free ligand. The  
6 latter was obtained by performing a quick simulation, involving both sampling and post-processing  
7 stages, on the sole ligand (absence of the protein binding site). The VS protocol is described in Fig.  
8  
9  
10  
11  
12  
13 2.C.

### 14 15 16 **Compound selection**

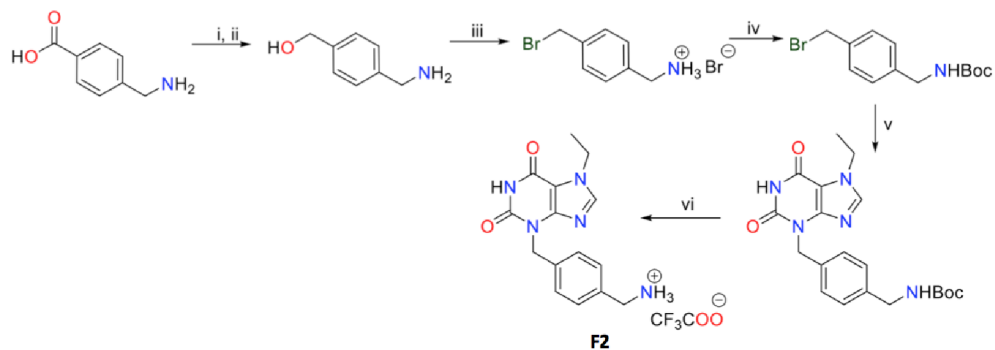
17  
18 The best poses within an energy window of 1 kcal/mol from the top 100 compounds were  
19 considered for further analysis (this threshold is a default value but can be user-defined). In addition,  
20 compounds that did not converge towards the targeted pocket area or compounds exhibiting a shift  
21 larger than 1.0 Å on the sole xanthine moiety were not considered. The shift was measured using the  
22 RMSD metric on the xanthine moiety between its reference constrained position and its final post-  
23 minimization position (after all constraints were removed). Remaining compounds were subjected to  
24 a clustering strategy based on chemical structures. In practice, compounds were simplified by  
25 automatically removing non-cyclic hydrophobic tails (carbon and halogen atoms), and clusters were  
26 generated by identifying identical compounds. For example, several compounds were diphenyl-  
27 analogs of compound **1** which was ranked in sixth position in the virtual screening, while its lowest-  
28 energy analog was ranked in first position. Similarly, 3 compounds were very close analogs of  
29 molecule **3** (the lowest-energy being ranked in position 5). Visual inspection of all poses was  
30 performed with Pymol (<http://pymol.org/>).

### 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 **Chemical synthesis**

47  
48 Commercially available reagents were used without additional purification. Automated syntheses  
49 were performed by using a Robotic platform Accelerator Synthesizer SLTII from Chemspeed  
50 technologies. Column chromatography was performed using Macherey-Nagel Kieselgel 60 (70-230  
51 mesh). The petroleum spirit refers to the fraction with distillation range 40-70°C. <sup>1</sup>H and <sup>13</sup>C NMR  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 spectra were recorded at room temperature in DMSO-*d*<sub>6</sub>, CD<sub>3</sub>OD or CDCl<sub>3</sub> by using a Bruker AC400,  
4  
5 AC250 or Agilent DD2 400 spectrometers. Chemical shifts ( $\delta$ ) are reported in parts per million (ppm)  
6  
7 with internal reference TMS and coupling values (J) in hertz. Abbreviations for peaks are, br: broad, s:  
8  
9 singlet, d: doublet, t: triplet, q: quadruplet, quint: quintuplet, sex: sextuplet and m: multiplet. The  
10  
11 spectra recorded are consistent with the proposed structures. Reaction monitoring and purity of  
12  
13 compounds were recorded by using analytical Agilent Infinity high performance liquid  
14  
15 chromatography (Column Zorbax SB-C18 1.8  $\mu$ M (2.1x50 mm); Mobile phase (A: 0.1% FA H<sub>2</sub>O, B: 0.1%  
16  
17 FA MeCN, Time/%B 0/10, 4/90, 7/90, 9/10, 10/10); Flow rate 0.3 mL/min; Diluent MeOH) with DAD  
18  
19 at 230 nM. All tested compounds yielded data consistent with a purity of  $\geq$  95%.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Scheme 1

Scheme 1. Synthesis of 3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt **F2**<sup>a</sup>

<sup>a</sup>reagents and conditions: i) LiAlH<sub>4</sub>, THF, 0°C-reflux, 12 hours; ii) H<sub>2</sub>O, NaOH, 0°C, 1h; iii) HBr, H<sub>2</sub>O, reflux, 3.5 hours; iv) Boc<sub>2</sub>O, NaHCO<sub>3</sub>, H<sub>2</sub>O-dioxane, rt, 3 hours; v) 7-ethyl-1H-purine-2,6(3H,7H)-dione, K<sub>2</sub>CO<sub>3</sub>, rt, 2 days; vi) CF<sub>3</sub>COOH, CH<sub>2</sub>Cl<sub>2</sub>, rt, 1 hour.

**(4-(aminomethyl)phenyl)methanol.** To a stirred suspension of commercially available 4-(aminomethyl)benzoic acid (10g, 66.2mmol) in tetrahydrofuran (100mL) lithium aluminium hydride (10g, 264.8mmol) was added portionwise at 0°C. The mixture was heated to reflux and stirred overnight before cooling down again to 0°C. 10mL of H<sub>2</sub>O were added, then 15mL of 10% NaOH and 30mL of H<sub>2</sub>O. The mixture was kept at 0°C for 1hour. Then the mixture was filtered through a pad of Celite and washed with EtOAc. The filtrate was concentrated to afford 4-(aminomethyl)benzyl alcohol as a yellow oil that crystallized to a beige solid after several minutes. The solid obtained was washed with Et<sub>2</sub>O to afford 4-(aminomethyl)benzyl alcohol **bb** (8.8g, yield 97%) as a light-beige powder. <sup>1</sup>H NMR (250 MHz, CD<sub>3</sub>OD) δ 7.32 (s, 4H), 4.58 (s, 2H), 3.78 (s, 2H). <sup>13</sup>C NMR (63 MHz, CD<sub>3</sub>OD) δ 142.2, 141.5, 128.5, 128.3, 65.0, 46.3.

**(4-(bromomethyl)phenyl)methanamine hydrobromic salt.** 4-(Aminomethyl)benzyl alcohol (768mg, 5.6mmol) was dissolved in a mixture of H<sub>2</sub>O (8.5mL) and HBr (46% aqueous solution, 13mL). The resulting mixture was refluxed for 3.5 hours, then the solvent was removed under reduced pressure,

1  
2  
3 and the solid obtained was washed with Et<sub>2</sub>O to afford 4-(bromomethyl)phenylmethanamine  
4 hydrobromic salt as a grey powder (1g, 89%). <sup>1</sup>H NMR (250 MHz, DMSO-*d*<sub>6</sub>) δ 8.26 (brs, 2H), 7.48 (s,  
5 4H), 4.71 (d, *J* = 3.6 Hz, 2H), 4.03 (s, 2H). <sup>13</sup>C NMR (63 MHz, DMSO-*d*<sub>6</sub>) δ 138.5, 134.1, 129.6, 129.4,  
6 41.9, 34.0.  
7  
8  
9

10  
11 **tert-butyl 4-(bromomethyl)benzylcarbamate.** To the solution of (4-  
12 (bromomethyl)phenyl)methanamine hydrobromic salt (1g, 5mmol) in a mixture of H<sub>2</sub>O and dioxane  
13 (40mL, 1:1), di-*tert*-butyl dicarbonate (3.3g, 15mmol) was added at 0°C. Then NaHCO<sub>3</sub> (0.87g,  
14 10mmol) was added at the same temperature, and the resulting mixture was stirred for 3 hours at  
15 room temperature. The mixture was extracted with Et<sub>2</sub>O (x3), combined organic layers were dried  
16 over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The resulting colorless oil was washed with  
17 petroleum ether and the precipitate formed was filtered off and dried under reduced pressure to  
18 afford *tert*-butyl 4-(bromomethyl)benzylcarbamate (1.1g, 73%) as a white solid. <sup>1</sup>H NMR (250 MHz,  
19 DMSO-*d*<sub>6</sub>) δ 7.38 (d, *J* = 7.5 Hz, 2H), 7.21 (d, *J* = 8.0 Hz, 2H), 4.68 (s, 2H), 4.10 (d, *J* = 6.1 Hz, 2H), 1.39  
20 (s, 9H). <sup>13</sup>C NMR (63 MHz, DMSO-*d*<sub>6</sub>) δ 155.8, 140.5, 136.4, 129.3, 127.2, 77.9, 43.1, 34.6, 28.3.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 **tert-butyl 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate.** To a suspension  
34 of 7-ethyl-1H-purine-2,6(3H,7H)-dione (180mg, 1mmol) prepared according to the previously  
35 described procedures<sup>27</sup> and K<sub>2</sub>CO<sub>3</sub> (151.8mg, 1.1mmol) in dimethylformamide (15mL) *tert*-butyl 4-  
36 (bromomethyl)benzylcarbamate (300mg, 1mmol) was added portionwise, and the resulting mixture  
37 was stirred at room temperature overnight. The solvent was removed under reduced pressure, and a  
38 crude product was purified by column chromatography (eluent 40:1:0.1 CH<sub>2</sub>Cl<sub>2</sub>:MeOH:NH<sub>3</sub>) to afford  
39 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate (202 mg, 51%) as a white  
40 solid. <sup>1</sup>H NMR (250 MHz, DMSO-*d*<sub>6</sub>) δ 8.07 (s, 1H), 7.30 (d, *J* = 8.1 Hz, 2H), 7.18 (d, *J* = 8.1 Hz, 2H), 5.09  
41 (s, 2H), 4.25 (q, *J* = 7.1 Hz, 2H), 4.09 (d, *J* = 6.0 Hz, 2H), 1.47-1.31 (m, 12H). <sup>13</sup>C NMR (63 MHz, DMSO-  
42 *d*<sub>6</sub>) δ 155.8, 154.7, 150.8, 149.7, 142.0, 139.4, 135.4, 127.6, 127.0, 106.5, 77.8, 44.5, 43.1, 41.6, 28.3,  
43 16.3.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt F2.** A suspension  
4 of 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate (200mg, 0.50mmol) in  
5 dichloromethane (3mL) was treated with trifluoroacetic acid (450 $\mu$ l), and the resulting mixture was  
6 stirred at room temperature for 1 hour. The solvent was removed under reduced pressure, and a  
7 crude product was purified by column chromatography (eluent 6:1:0.1, CH<sub>2</sub>Cl<sub>2</sub>:MeOH:NH<sub>3</sub>) to afford  
8 3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt **F2** (207mg, 100%)  
9 as a white solid. <sup>1</sup>H NMR (250 MHz, CD<sub>3</sub>OD)  $\delta$  7.96 (s, 1H), 7.51 (d, *J* = 8.2 Hz, 2H), 7.42 (d, *J* = 8.2 Hz,  
10 2H), 5.26 (s, 2H), 4.36 (q, *J* = 7.2 Hz, 2H), 4.10 (s, 2H), 1.51 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (63 MHz,  
11 CD<sub>3</sub>OD)  $\delta$  156.5, 152.9, 151.6, 143.2, 139.2, 134.0, 130.2, 129.8, 108.5, 46.2, 44.0, 43.4, 16.8.

12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23 **General procedure for the synthesis of N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-**  
24 **yl)methyl)benzyl)sulfonamide derivatives.** The automated sulfonylation were run in 2 mL LC glass  
25 vials. A stock solution of 3-(4-(aminomethyl)benzyl)-7-ethylxanthine trifluoroacetic salt **F2** (206.5 mg,  
26 1eq) and triethylamine (138.7  $\mu$ L, 2eq) in anhydrous tetrahydrofuran (20 mL) was prepared and  
27 placed on the robotic deck. The robot automatically prepared the reaction mixtures by dispensing  
28 the prepared solution (400  $\mu$ L) to the individual vials containing chlorosulfonyl derivatives (0.01  
29 mmol) via the liquid handling tool. After 12 hours at room temperature, samples were taken out and  
30 analyzed by LCMS to determine the percent of conversion. The product solutions were concentrated  
31 at 80°C, and residues were diluted with dimethylsulfoxide to afford a solution 5.10<sup>-2</sup> M of the  
32 expected compounds. Further LCMS analysis allowed rectifying the final concentration of the  
33 products precisely to 10<sup>-2</sup> M for further biological evaluations.

34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-4-sulfonamide (1)**

47  
48  
49 LCMS C<sub>27</sub>H<sub>25</sub>N<sub>5</sub>O<sub>4</sub>S Rt = 6.691, m/z found 515.5, calcd 515.6.

50  
51 **4-(benzyloxy)-N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzene-**

52  
53 **sulfonamide (2)** LCMS C<sub>28</sub>H<sub>27</sub>N<sub>5</sub>O<sub>5</sub>S Rt = 6.659, m/z found 545.5, calcd 545.6.

1  
2  
3 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-(2-methyloxazol-4-yl)benzene-**  
4 **sulfonamide (3)** LCMS  $C_{25}H_{24}N_6O_5S$  Rt = 5.998, m/z found 520.5, calcd 520.6.

6  
7 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzo[b]thiophene-2-sulfonamide**  
8 **(4)** LCMS  $C_{23}H_{21}N_5O_4S_2$  Rt = 6.443, m/z found 495.6, calcd 495.6.

10  
11 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-5-(3-methyl-1H-pyrazol-5-**  
12 **yl)thiophene-2-sulfonamide (5)** LCMS  $C_{23}H_{23}N_7O_4S_2$  Rt = 5.901, m/z found 525.5, calcd 525.6.

14  
15 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methoxybenzene-sulfonamide**  
16 **(6)** LCMS  $C_{22}H_{23}N_5O_5S$  Rt = 5.998, m/z found 469.6, calcd 469.5.

18  
19 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-phenoxybenzene-sulfonamide**  
20 **(7)** LCMS  $C_{27}H_{25}N_5O_5S$  Rt = 6.611, m/z found 531.5, calcd 531.6.

22  
23 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1,3-dimethyl-2-oxo-2,3-dihydro-**  
24 **1H-benzo[d]imidazole-5-sulfonamide (8)** LCMS  $C_{24}H_{25}N_7O_5S$  Rt = 5.579, m/z found 523.5, calcd 523.6.

26  
27 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-oxo-1,2,3,4-**  
28 **tetrahydroquinoline-6-sulfonamide (9)** LCMS  $C_{24}H_{24}N_6O_5S$  Rt = 5.353, m/z found 508.5, calcd 508.6.

30  
31 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1-phenyl-1H-pyrazole-3-**  
32 **sulfonamide (10)** LCMS  $C_{24}H_{23}N_7O_4S$  Rt = 6.207, m/z found 505.5, calcd 505.6.

34  
35 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-phenylethane-sulfonamide (11)**  
36 **(11)** LCMS  $C_{23}H_{25}N_5O_4S$  Rt = 6.111, m/z found 467.6, calcd 467.5.

38  
39 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzo[d]isoxazole-5-sulfonamide**  
40 **(12)** LCMS  $C_{22}H_{20}N_6O_5S$  Rt = 5.530, m/z found 480.6, calcd 480.5.

42  
43 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methylbenzene-sulfonamide**  
44 **(13)** LCMS  $C_{22}H_{23}N_5O_4S$  Rt = 6.143, m/z found 453.6, calcd 453.5.

46  
47 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-3-sulfonamide**  
48 **(14)** LCMS  $C_{27}H_{25}N_5O_4S$  Rt = 6.611, m/z found 515.5, calcd 515.6.

**N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-morpholinoethane-sulfonamide**

**(15)** LCMS  $C_{21}H_{28}N_6O_5S$  Rt = 5.627, m/z found 476.6, calcd 476.6.

**N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-5-(isoxazol-5-yl)-2-**

**methylbenzene-sulfonamide (16)** LCMS  $C_{25}H_{24}N_6O_5S$  Rt = 6.030, m/z found 520.5, calcd 520.6.

**N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-oxo-2,3,4,5-tetrahydro-1H-**

**benzo[b]azepine-7-sulfonamide (17)** LCMS  $C_{25}H_{26}N_6O_5S$ , Rt = 5.370, m/z found 522.5, calcd 522.6.

**General procedure for the purification of N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)sulfonamide derivatives.** According to the procedure described above derivatives **1-10, 13, 14, 17-19** were prepared on a larger amount of starting materials. The solvent was removed under reduced pressure, and crude products were purified by column chromatography to afford desired compounds with purity >95%.

**N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-4-sulfonamide (1)**

(eluent  $CH_2Cl_2$ :MeOH 10:0.5, yield 89%) as light yellow solid.  $^1H$  NMR (400 MHz, DMSO)  $\delta$  11.20 (s), 8.14 (t,  $J$  = 6.3 Hz), 8.02 (s), 7.86 (s), 7.76 – 7.71 (m), 7.52 (dd,  $J$  = 10.3, 4.7 Hz), 7.47 – 7.41 (m), 7.25 (d,  $J$  = 8.2 Hz), 7.19 (d,  $J$  = 8.1 Hz), 5.05 (s), 4.22 (q,  $J$  = 7.2 Hz), 3.96 (d,  $J$  = 6.2 Hz), 1.37 (t,  $J$  = 7.2 Hz).

$^{13}C$  NMR (101 MHz, DMSO)  $\delta$  154.59, 150.76, 149.66, 143.90, 141.90, 139.39, 138.59, 136.76, 136.00, 129.12, 128.45, 127.70, 127.53, 127.39, 127.17, 127.07, 106.44, 45.88, 44.50, 41.52, 16.18. LCMS  $C_{27}H_{25}N_5O_4S$ , Rt = 6.691, m/z found 515.5, calcd 515.6.

**4-(benzyloxy)-N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzene-**

**sulfonamide (2)** (eluent  $CH_2Cl_2$ :MeOH 10:0.5, yield 83%) as white solid.  $^1H$  NMR (400 MHz, DMSO)  $\delta$  11.21 (s), 8.03 (s), 7.92 (t,  $J$  = 6.3 Hz), 7.72 (d,  $J$  = 8.8 Hz), 7.47 (d,  $J$  = 7.1 Hz), 7.41 (t,  $J$  = 7.3 Hz), 7.36 (d,  $J$  = 7.1 Hz), 7.24 (d,  $J$  = 8.1 Hz), 7.20 – 7.11 (m), 5.20 (s), 5.05 (s), 4.22 (q,  $J$  = 7.1 Hz), 3.88 (d,  $J$  = 6.2 Hz), 1.37 (t,  $J$  = 7.2 Hz).  $^{13}C$  NMR (101 MHz, DMSO)  $\delta$  161.17, 154.59, 150.77, 149.67, 141.91, 136.81, 136.37, 135.95, 132.46, 128.64, 128.50, 128.05, 127.82, 127.68, 127.51, 115.09, 106.45, 69.62, 45.84, 44.50, 41.52, 16.18. LCMS  $C_{28}H_{27}N_5O_5S$ , Rt = 6.659, m/z found 545.5, calcd 545.6.

1  
2  
3 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-(2-methyloxazol-4-yl)benzene-**  
4 **sulfonamide (3)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 90%) as white solid. <sup>1</sup>H NMR (250 MHz, DMSO) δ  
5 11.25 (s, 1H), 8.65 (s, 1H), 8.13 (t, *J* = 6.1 Hz, 1H), 8.05 (s, 1H), 7.94 (d, *J* = 8.3 Hz, 2H), 7.83 (d, *J* = 8.3  
6 Hz, 2H), 7.22 (q, *J* = 8.2 Hz, 4H), 5.05 (s, 2H), 4.23 (q, *J* = 7.1 Hz, 2H), 3.94 (d, *J* = 6.0 Hz, 2H), 2.50 (s,  
7 3H), 1.38 (t, *J* = 7.1 Hz, 3H). <sup>13</sup>C NMR (63 MHz, DMSO) δ 162.1, 154.7, 150.8, 149.7, 142.0, 139.4,  
8 138.5, 136.8, 136.6, 136.1, 134.8, 127.8, 127.6, 127.2, 125.6, 106.5, 45.9, 44.6, 41.6, 16.2, 13.6. LCMS  
9 C<sub>25</sub>H<sub>24</sub>N<sub>6</sub>O<sub>5</sub>S, Rt = 6.014, m/z found 521.0, calcd 520.6.

10  
11  
12  
13  
14  
15  
16  
17 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzo[b]thiophene-2-sulfonamide**  
18 **(4)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 76%) as light yellow plates. <sup>1</sup>H NMR (400 MHz, DMSO) δ 11.20  
19 (s, 1H), 8.55 (t, *J* = 6.1 Hz, 1H), 8.07 (d, *J* = 7.8 Hz, 1H), 8.03 (s, 1H), 8.01 – 7.96 (m, 1H), 7.94 (s, 1H),  
20 7.59 – 7.42 (m, 2H), 7.22 (q, *J* = 8.4 Hz, 4H), 5.03 (s, 2H), 4.22 (q, *J* = 7.1 Hz, 2H), 4.08 (d, *J* = 5.8 Hz,  
21 2H), 1.38 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (101 MHz, DMSO) δ 154.58, 150.73, 149.63, 141.89, 141.78,  
22 140.63, 137.59, 136.43, 136.07, 128.53, 127.69, 127.59, 127.07, 125.71, 125.42, 122.98, 106.43,  
23 46.04, 44.46, 41.51, 16.17. LCMS C<sub>23</sub>H<sub>21</sub>N<sub>5</sub>O<sub>4</sub>S<sub>2</sub>, Rt = 6.443, m/z found 495.6, calcd 495.6.

24  
25  
26  
27  
28  
29  
30  
31  
32 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-5-(3-methyl-1H-pyrazol-5-**  
33 **yl)thiophene-2-sulfonamide (5)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:1, yield 66%) as white solid. <sup>1</sup>H NMR (400  
34 MHz, DMSO) δ 11.20 (s), 8.29 (t, *J* = 6.1 Hz), 8.03 (s), 7.47 (d, *J* = 3.8 Hz), 7.31 (d, *J* = 3.7 Hz), 7.26 (d, *J*  
35 = 7.8 Hz), 7.21 (d, *J* = 7.8 Hz), 6.46 (s), 5.05 (s), 4.22 (q, *J* = 7.1 Hz), 4.02 (d, *J* = 5.9 Hz), 2.26 (s), 1.37 (t,  
36 *J* = 7.1 Hz). <sup>13</sup>C NMR (101 MHz, DMSO) δ 154.59, 150.76, 149.66, 141.91, 138.41, 136.60, 136.05,  
37 132.12, 127.70, 127.58, 123.14, 106.45, 101.73, 46.04, 44.51, 41.52, 16.18, 10.48. LCMS  
38 C<sub>23</sub>H<sub>23</sub>N<sub>7</sub>O<sub>4</sub>S<sub>2</sub>, Rt = 5.901, m/z found 525.5 calcd 525.6.

39  
40  
41  
42  
43  
44  
45  
46  
47 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methoxybenzene-sulfonamide**  
48 **(6)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 71%) as white solid. <sup>1</sup>H NMR (250 MHz, CDCl<sub>3</sub>) δ 7.65 – 7.56 (m,  
49 2H), 7.53 (s, 1H), 7.20 (d, *J* = 8.1 Hz, 2H), 7.00 (d, *J* = 8.1 Hz, 2H), 6.86 – 6.76 (m, 2H), 5.02 (s, 2H), 4.17  
50 (d, *J* = 7.2 Hz, 2H), 3.86 (s, 2H), 3.72 (s, 3H), 1.36 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (63 MHz, CDCl<sub>3</sub>) δ 162.7,  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 154.9, 151.2, 150.1, 140.9, 136.3, 135.4, 131.4, 128.9, 128.3, 127.9, 114.1, 107.4, 55.4, 46.4, 45.3,  
4  
5 42.4, 16.1. **LCMS** C<sub>22</sub>H<sub>23</sub>N<sub>5</sub>O<sub>5</sub>S, Rt = 5.822, m/z found 469.9, calcd 496.5.

6  
7 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-phenoxybenzene-sulfonamide**

8  
9 **(7)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 86%) as white solid. <sup>1</sup>H NMR (250 MHz, CDCl<sub>3</sub>+CD<sub>3</sub>OD) δ 7.71  
10  
11 (dd, *J* = 6.6, 2.2 Hz, 3H), 7.44 – 7.29 (m, 4H), 7.17 (dd, *J* = 18.9, 7.6 Hz, 3H), 7.07 – 6.90 (m, 4H), 5.14  
12  
13 (s, 2H), 4.26 (q, *J* = 7.1 Hz, 2H), 4.01 (s, 2H), 1.46 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (63 MHz, CDCl<sub>3</sub>+CD<sub>3</sub>OD) δ  
14  
15 161.88, 155.66, 155.61, 151.91, 150.70, 141.76, 137.03, 136.02, 134.30, 130.60, 129.56, 128.80,  
16  
17 128.45, 125.35, 120.65, 117.93, 107.94, 46.97, 45.86, 42.91, 16.57. **LCMS** C<sub>27</sub>H<sub>25</sub>N<sub>5</sub>O<sub>5</sub>S, Rt = 6.611,  
18  
19 m/z found 531.5, calcd 531.6.

20  
21  
22 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1,3-dimethyl-2-oxo-2,3-dihydro-**

23  
24 **1H-benzo[d]imidazole-5-sulfonamide (8)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 100%) as white solid. <sup>1</sup>H  
25  
26 **NMR** (250 MHz, DMSO) δ 11.23 (s, 1H), 8.03 (s, 1H), 7.95 (t, *J* = 6.3 Hz, 1H), 7.53 (d, *J* = 8.2 Hz, 1H),  
27  
28 7.48 (s, 1H), 7.30 (d, *J* = 8.2 Hz, 1H), 7.19 (q, *J* = 8.1 Hz, 4H), 5.02 (s, 2H), 4.22 (dd, *J* = 14.5, 7.3 Hz, 2H),  
29  
30 3.89 (d, *J* = 6.2 Hz, 2H), 3.35 (s, 3H), 3.16 (d, *J* = 5.2 Hz, 3H), 1.37 (t, *J* = 7.1 Hz, 3H). <sup>13</sup>C NMR (63 MHz,  
31  
32 DMSO) δ 154.7, 154.2, 150.8, 149.7, 142.0, 136.9, 136.0, 132.9, 132.7, 129.6, 127.8, 127.5, 120.2,  
33  
34 107.6, 106.5, 106.1, 55.0, 48.7, 41.6, 27.3, 27.2, 16.3. **LCMS** C<sub>24</sub>H<sub>25</sub>N<sub>7</sub>O<sub>5</sub>S, Rt = 5.515, m/z found 524.0,  
35  
36 calcd 523.6.

37  
38  
39 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-oxo-1,2,3,4-**

40  
41 **tetrahydroquinoline-6-sulfonamide (9)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:1, yield 93%) as white solid. <sup>1</sup>H NMR  
42  
43 (400 MHz, DMSO) δ 11.20 (s), 10.41 (s), 8.03 (s), 7.91 (t, *J* = 6.3 Hz), 7.61 – 7.54 (m), 7.24 (d, *J* = 8.2  
44  
45 Hz), 7.17 (d, *J* = 8.1 Hz), 6.96 (d, *J* = 8.1 Hz), 5.05 (s), 4.22 (q, *J* = 7.1 Hz), 3.90 (d, *J* = 6.2 Hz), 2.91 (t, *J* =  
46  
47 7.6 Hz), 2.49 – 2.45 (m), 1.38 (t, *J* = 7.2 Hz). <sup>13</sup>C NMR (101 MHz, DMSO) δ 170.30, 154.57, 150.75,  
48  
49 149.65, 141.90, 141.79, 136.86, 135.90, 133.45, 127.68, 127.46, 126.27, 126.09, 124.05, 114.97,  
50  
51 106.43, 45.82, 44.49, 41.52, 29.85, 24.46, 16.17. **LCMS** C<sub>24</sub>H<sub>24</sub>N<sub>6</sub>O<sub>5</sub>S, Rt = 5.353, m/z found 508.5,  
52  
53 calcd 508.6.

1  
2  
3 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1-phenyl-1H-pyrazole-3-**  
4 **sulfonamide (10)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.7, yield 60%) as light orange solid. <sup>1</sup>H NMR (400 MHz,  
5 DMSO) δ 11.20 (s), 8.96 (s), 8.02 (d, *J* = 9.3 Hz), 7.88 (d, *J* = 7.8 Hz), 7.54 (t, *J* = 7.9 Hz), 7.40 (t, *J* = 7.4  
6 Hz), 7.25 (q, *J* = 8.3 Hz), 5.02 (s), 4.22 (q, *J* = 7.2 Hz), 4.05 (d, *J* = 6.2 Hz), 1.37 (t, *J* = 7.2 Hz). <sup>13</sup>C NMR  
7 (101 MHz, DMSO) δ 154.58, 150.75, 149.65, 141.90, 139.13, 138.75, 136.81, 135.97, 129.64, 129.29,  
8 127.75, 127.58, 127.51, 124.86, 119.17, 106.43, 45.88, 44.47, 41.51, 16.17. LCMS C<sub>24</sub>H<sub>23</sub>N<sub>7</sub>O<sub>4</sub>S, Rt =  
9 6.207, m/z found 505.5, calcd 505.6.  
10  
11  
12  
13  
14  
15  
16  
17

18 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methylbenzene-sulfonamide**  
19 **(13)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 79%) as white solid. <sup>1</sup>H NMR (250 MHz, CD<sub>3</sub>OD+CDCl<sub>3</sub>) δ 7.81  
20 (s, 1H), 7.66 (d, *J* = 8.3 Hz, 2H), 7.29 (dd, *J* = 8.0, 6.4 Hz, 4H), 7.13 (d, *J* = 8.3 Hz, 2H), 5.14 (s, 2H), 4.31  
21 (q, *J* = 7.2 Hz, 2H), 3.98 (s, 2H), 2.39 (s, 3H), 1.48 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (63 MHz, CD<sub>3</sub>OD+CDCl<sub>3</sub>) δ  
22 156.0, 152.3, 151.1, 144.1, 142.4, 138.2, 137.5, 136.4, 130.3, 128.9, 128.7, 127.6, 108.2, 47.2, 46.0,  
23 43.1, 21.6, 16.7. LCMS C<sub>22</sub>H<sub>23</sub>N<sub>5</sub>O<sub>4</sub>S, Rt = 5.983, m/z found 453.9, calcd 453.5.  
24  
25  
26  
27  
28  
29  
30

31 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-3-sulfonamide**  
32 **(14)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:0.5, yield 96%) as a white solid. <sup>1</sup>H NMR (400 MHz, DMSO) δ 11.20 (s,  
33 1H), 8.17 (t, *J* = 6.3 Hz, 1H), 8.03 (s, 1H), 7.98 (t, *J* = 1.7 Hz, 1H), 7.86 – 7.80 (m, 1H), 7.75 – 7.70 (m,  
34 1H), 7.65 (dd, *J* = 5.2, 3.3 Hz, 2H), 7.59 (t, *J* = 7.8 Hz, 1H), 7.53 – 7.47 (m, 2H), 7.43 (dt, *J* = 9.4, 4.2 Hz,  
35 1H), 7.21 (d, *J* = 8.3 Hz, 2H), 7.15 (d, *J* = 8.2 Hz, 2H), 5.01 (s, 2H), 4.22 (q, *J* = 7.1 Hz, 2H), 3.99 (d, *J* = 6.2  
36 Hz, 2H), 1.37 (t, *J* = 7.2 Hz, 3H). <sup>13</sup>C NMR (101 MHz, DMSO) δ 154.57, 150.72, 149.64, 141.87, 141.45,  
37 141.05, 138.67, 136.56, 135.92, 130.40, 129.76, 129.10, 128.18, 127.69, 127.58, 127.52, 126.82,  
38 125.29, 124.46, 106.41, 45.94, 44.46, 41.52, 16.18. LCMS C<sub>27</sub>H<sub>25</sub>N<sub>5</sub>O<sub>4</sub>S, Rt = 6.611, m/z found 515.5,  
39 calcd 515.6.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 **N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-oxo-2,3,4,5-tetrahydro-1H-**  
51 **benzo[b]azepine-7-sulfonamide (17)** (eluent CH<sub>2</sub>Cl<sub>2</sub>:MeOH 10:1, yield 69%) as a white solid. <sup>1</sup>H NMR  
52 (250 MHz, DMSO-*d*<sub>6</sub>) δ 9.83 (s, 1H), 8.03 (s, 1H), 7.65 – 7.58 (m, *J* = 7.1 Hz, 2H), 7.22 (d, *J* = 8.1 Hz,  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 2H), 7.14 (d, J = 8.1 Hz, 2H), 7.07 (d, J = 9.0 Hz, 1H), 5.03 (s, 2H), 4.22 (q, J = 7.2 Hz, 2H), 3.95 (s, 2H),  
4  
5 2.70 (t, J = 6.6 Hz, 2H), 2.24 – 2.01 (m, 4H), 1.37 (t, J = 7.1 Hz, 3H); <sup>13</sup>C NMR (63 MHz, DMSO-*d*<sub>6</sub>) δ  
6  
7 173.2, 154.6, 150.8, 149.7, 142.7, 142.0, 136.8, 136.2, 135.9, 134.1, 128.1, 127.7(x2), 127.5(x2),  
8  
9 125.8, 121.6, 106.5, 45.9, 44.5, 41.6, 33.0, 30.0, 27.7, 16.2. LCMS C<sub>25</sub>H<sub>26</sub>N<sub>6</sub>O<sub>5</sub>S, Rt = 5.370, m/z found  
10  
11 522.5, calcd 522.6.

### 12 13 14 **Robotic platform Labcyte workstation**

15  
16 The robotic workstation “Access Labcyte” is a secured enclosure articulated around a robotic arm, a  
17  
18 nanovolume dispenser “Echo®550” from Labcyte and a microvolume dispenser “Multidrop Combi”  
19  
20 from Thermo Fisher Scientific. The system is also composed of a microplate centrifuge “Velocity 11”,  
21  
22 a microplate thermal sealer under argon “PlateLoc” from Agilent Technologies and an absorbance  
23  
24 and fluorescence reader “PHERAstar FS” from BMG Labtech. The “Echo®550” is based on the  
25  
26 acoustic liquid handling technology that uses acoustic waves to automatically and very accurately  
27  
28 dispense nanovolumes (2.5 nL droplets) without contact, therefore, eliminating risks of cross-  
29  
30 contamination or sample absorption on tips.  
31

### 32 33 **Protein expression and purification**

34  
35 For Homogeneous Time Resolved Fluorescence (HTRF) experiments, a BRD4(BD1) synthetic gene that  
36  
37 includes a TEV cleavage site was purchased from LifeTechnology in a pDONR transport vector before  
38  
39 cloning into a pDEST15 expression vector for GST affinity purification. Protein production was carried  
40  
41 out using similar protocols as used for the His-BRD4(BD1) system. Purification was carried on GST  
42  
43 affinity resin (Thermo Scientific) and reduced glutathione was used for protein release. GST-  
44  
45 BRD4(BD1) was further purified by size exclusion chromatography on a Superdex 16/60 Hiload  
46  
47 column (GE Healthcare) using 50 mM HEPES pH7.5, 150 mM NaCl Buffer.  
48  
49

50  
51 For Isothermal Titration Calorimetry (ITC), Thermal Shift Assay (TSA) and crystallogenesis, BRD4(BD1)  
52  
53 was produced and purified using a histidine tag affinity chromatography as described by  
54  
55 Filipakopoulos *et al.*<sup>50</sup> For these experiments, a pNIC28-BSA4 expression vector containing  
56  
57  
58  
59  
60

1  
2  
3 BRD4(BD1) and a Tobacco Etch Virus (TEV) protease cleavage site have kindly been provided by  
4 Stefan Knapp laboratory from the SGC at the University of Oxford. After size exclusion  
5 chromatography, the fractions presenting pure BRD4(BD1) after TEV cleavage of the histidine tag  
6 were pooled and concentrated to 25 mg/mL for crystallogensis. For ITC and TSA assays, the protein  
7 was concentrated up to at 6 mg/mL and the DTT was removed using a buffer exchange column (PD10  
8 from GE healthcare) equilibrated with 10 mM HEPES pH7.5, 150 mM NaCl.

### 15 **HTRF assay**

16  
17  
18 All HTRF reagents (donor and acceptor) were purchased from Cisbio and reconstituted according to  
19 the supplier protocols. The protein BRD4(BD1) was produced and purified in house. The peptide  
20 H4KAc5/8/12/16 (SGRGK(Ac)CGK(Ac)GLGK(Ac)GGAK(Ac)RHRKVG), was purchased from Genic Bio  
21 Synthetic. The IC<sub>50</sub> measurements (10 points) were performed in triplicates and the concentration  
22 range was comprised between 50 μM and 13 nM. HTRF assays were performed in white 384 Well  
23 Small Volume™ HiBase Polystyrene Microplates (Greiner bio-one) with a total working volume of 20  
24 μL. Compounds and DMSO were stored in 384-LDV plates (Labcyte) and reagents (protein, donor and  
25 acceptor) in 384-PP plates (Labcyte). 100 nL of compounds or DMSO (for positive and negative  
26 controls at 0.5% final DMSO concentration) were dispensed from a concentration stock of 10 mM in  
27 100% DMSO using an Echo®550 (Labcyte). For each assay, 200 nL of master mix (protein + donor +  
28 acceptor) and 200 nL of peptide were added in the assay wells. The final volume was completed with  
29 19.5 μL of buffer (50 mM HEPES at pH 7.5, 400 mM KF, 0.1% BSA), using a Multidrop Combi (Thermo  
30 Fisher Scientific).

31  
32  
33 HTRF signals were measured, after a final incubation (3h at room temperature), using a PHERAstar FS  
34 (BMG Labtech) with an excitation filter at 337 nm and fluorescence wavelength measurement at 620  
35 and 665 nm (60 μs integration delay and 400 μs integration time). Percentage of inhibition was  
36 calculated using the following equation: % inhibition = [(compound signal) - (min signal)] / [(max  
37 signal) - (min signal)] \* 100, where “max signal” is the signal ratio ([intensity (665 nm)/intensity (620  
38 nm)] / [intensity (665 nm)/intensity (620 nm)]).

1  
2  
3 nm)]\*10<sup>4</sup>) with the compound vehicle alone (DMSO) and “min signal” the signal ratio without  
4  
5 peptide. For IC<sub>50</sub> measurements, values were normalized and fitted with Prism 5.03 (GraphPad  
6  
7 software) using the following equation:  $Y = 100 / (1 + ((X / IC_{50})^{Hill\ slope}))$ .  
8

### 9 **Isothermal titration calorimetry**

10  
11 ITC was used to evaluate the thermodynamics parameters of the binding between BRD4(BD1) and  
12  
13 the selected compounds, using ITC conditions previously described by Filippakopoulos *et al.*<sup>50</sup>  
14  
15 Purified BRD4(BD1) was extensively dialyzed in the ITC buffer containing 10 mM HEPES pH7.5 and  
16  
17 150 mM NaCl. Compounds were diluted directly in the last protein dialysate buffer prior to  
18  
19 experiments. Titrations were carried out on a MicroCal ITC200 microcalorimeter (GE Healthcare,  
20  
21 Piscataway, NJ). Each experiment was designed as reverse titrations experiments (protein in the  
22  
23 syringe and ligand in the cell) using 13 injections at 15 or 25°C. A first small injection (generally 0.4  
24  
25 μL) was included in the titration protocol in order to remove air bubbles trapped in the syringe prior  
26  
27 to the titration. ITC experiments were performed in triplicates.  
28  
29

30  
31 Raw data were scaled after setting the zero to the titration saturation heat value. Integrated raw ITC  
32  
33 data were fitted to a one site non-linear least squares fit model using the MicroCal Origin plugin as  
34  
35 implemented in Origin 7 (Origin Lab). Finally, ΔG and TΔS values were calculated from the fitted ΔH  
36  
37 and K<sub>D</sub> values using the equations  $\Delta G = -R.T.\ln K_D$  and  $\Delta G = \Delta H - T\Delta S$ .  
38  
39

### 40 **Thermal shift assay**

41  
42 The TSA experiments were performed in triplicates, in 384-Well RT-PCR plates (Hard-Sheel® PCR  
43  
44 plate, 384-well, thin-wall from Bio-Rad). The reagents (compound, protein and fluorophore) were  
45  
46 dispensed using an Echo®550 (Labcyte). 100 nL of compound at a final concentration of 10 μM in  
47  
48 100% DMSO (final concentration 0.1% or 0.5% DMSO) were dispensed in the assay plate. 200 nL of  
49  
50 Thermal Shift™ Dye (Thermo Fisher Scientific) diluted at a final concentration of 1:1000 and 300 nL of  
51  
52 BRD4(BD1) at a final concentration of 2 μM were added in the assay wells and the volume was  
53  
54 completed at 19.5 μL with the assay buffer (10 mM HEPES pH 7.5, 500 mM NaCl) using a Multidrop  
55  
56  
57  
58  
59  
60

1  
2  
3 Combi (Thermo Fisher Scientific). The plates were sealed with optical film (Ampliseal, Greiner) and  
4  
5 centrifuged at 1000 rpm for 1 min at 4°C. The thermal melting experiments were carried using a  
6  
7 CFX384 RTPCR (Bio-Rad). The plates were first equilibrated at 25°C for 1 min, then plates were  
8  
9 heated from 25 to 70°C by 0.5°C steps (25 second equilibration). Raw fluorescence data were treated  
10  
11 and the melting temperatures ( $T_m$ ) were calculated using CFX Manager 3.1 (Bio-Rad).  
12

### 13 **Structural characterization**

14  
15  
16 BRD4(BD1)-inhibitor co-crystallization was performed at 19°C (292K) using the hanging drop vapor  
17  
18 diffusion method. The crystallization conditions are described in Table S2. The compounds and  
19  
20 BRD4(BD1) preparation were mixed at a 1:1 ratio with the precipitant solution and crystals grew to  
21  
22 diffracting quality within 7-15 days. Data were collected at the ESRF on the beamlines ID30A-1 for  
23  
24 fragment **F1** and ID29 for compound **17**. Indexing, integration and scaling were performed using  
25  
26 XDS.<sup>51, 52</sup> Initial phases were calculated by molecular replacement with Phaser MR from the CCP4  
27  
28 suite<sup>53</sup> using a 3D structure of the first domain of BRD4 protein extracted from the Protein Data Bank  
29  
30 (accession code: 2OSS). Initial models for the protein and ligands were built in COOT.<sup>54</sup> The cycles of  
31  
32 refinement were carried out with Refmac5.<sup>55</sup> Data collection and refinement statistics can be found  
33  
34 in Table S3.  
35

### 36 **Supporting Information**

37  
38  
39 Additional supplemental tables and figures are provided free of charge on the ACS publication  
40  
41 website.  
42

43  
44  
45  
46  
47 - Generic building of the focused virtual library; 2D structure of all compounds; X-ray structural data  
48  
49 for F1 and 17 bound to BRD4(BD1); ITC data of the purified compounds (PDF)  
50

51  
52  
53  
54 - Molecular formula strings and associated affinity data (CSV)  
55  
56  
57  
58  
59  
60

### Accession codes

Atomic coordinates and structure factors have been deposited into the Protein Data Bank (PDB) under accession codes 6FNX (**F1**) and 6FO5 (**17**). Authors will release the atomic coordinates and experimental data upon article publication.

### Author information

\*Correspondence should be addressed to [philippe.roche@inserm.fr](mailto:philippe.roche@inserm.fr) and [xavier.morelli@inserm.fr](mailto:xavier.morelli@inserm.fr)

<sup>1</sup>Laurent Hoffer and Yuliia Voitovich contributed equally to this work

### ORCID

[laurent.hoffer@inserm.fr](mailto:laurent.hoffer@inserm.fr), 0000-0003-1906-7128

[voitovich\\_25@mail.ru](mailto:voitovich_25@mail.ru), 0000-0002-9282-3834

[brigitt.raux@gmail.com](mailto:brigitt.raux@gmail.com), 0000-0002-2194-0259

[derviauxc@ipc.unicancer.fr](mailto:derviauxc@ipc.unicancer.fr), 0000-0001-7441-1126

[stephane.betzi@inserm.fr](mailto:stephane.betzi@inserm.fr), 0000-0001-5935-5058

[dhorvath@unistra.fr](mailto:dhorvath@unistra.fr), 0000-0003-0173-5714

[varnek@unistra.fr](mailto:varnek@unistra.fr), 0000-0003-1886-925X

[yves.collette@inserm.fr](mailto:yves.collette@inserm.fr), 0000-0001-5359-7099

[sebastien.combes@univ-amu.fr](mailto:sebastien.combes@univ-amu.fr), 0000-0002-8213-2407

[philippe.roche@inserm.fr](mailto:philippe.roche@inserm.fr), 0000-0002-5580-0588

[xavier.morelli@inserm.fr](mailto:xavier.morelli@inserm.fr), 0000-0001-8101-7901

### Acknowledgments

The project leading to this publication received funding from the Excellence Initiative of Aix-Marseille University A\*MIDEX, a French “Investissements d’Avenir” program. This study was partly supported by research funding from the Canceropôle PACA, Institut National du Cancer and Région Provence-Alpes-Côte d’Azur (Grant #2018-03), Fondation ARC (PJA20171206125) and ANR grant (ANR-15-CE18-

0023). LH was supported by a fellowship from A\*MIDEX; YV was supported by a Ph.D. fellowship “Metchnikov” from the French government and a fellowship from the “Fondation ARC pour la Recherche sur le Cancer”; BR was supported by a fellowship from the “Fondation pour la Recherche Médicale” (FRM). We would like to thank the FRISBI ANR-10-INSB-05-01 grant for access time on the structural biology platform of the AFMB laboratory. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities (beam time application MX1696 and MX1785) and the staff at beamlines ID29 and ID30A-1 (Nicolas Foos and Nurizzo Didier). We thank Bernard Chetrit from the Datacentre IT and Scientific Computing facility of the CRCM.

### Abbreviations used

BB, building blocks; BD1, first domain of bromodomain; BET, bromodomain and extra-terminal; BRD, bromodomain; DOTS, diversity-oriented target-focused synthesis; H2L, hit-to-lead; HTRF, homogeneous time-resolved fluorescence; ITC, isothermal titration calorimetry; Kac, acetylated lysine; TSA, thermal shift assay; VS, virtual screening.

### References

1. Erakovic Haber, V.; Spaventi, R. Discovery and development of novel drugs. *Prog Mol Subcell Biol* **2017**, *55*, 91-104.
2. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br J Pharmacol* **2011**, *162*, 1239-1249.
3. Wigglesworth, M. J.; Murray, D. C.; Blackett, C. J.; Kossenjans, M.; Nissink, J. W. Increasing the delivery of next generation therapeutics from high throughput screening libraries. *Curr Opin Chem Biol* **2015**, *26*, 104-110.
4. Romasanta, A. K. S.; van der Sijde, P.; Hellsten, I.; Hubbard, R. E.; Keseru, G. M.; van Muijlwijk-Koezen, J.; de Esch, I. J. P. When fragments link: a bibliometric perspective on the development of fragment-based drug discovery. *Drug Discov Today* **2018**, *18*, 30084-30089.
5. Ferreira, L. G.; Andricopulo, A. D. From protein structure to small-molecules: recent advances and applications to fragment-based drug discovery. *Curr Top Med Chem* **2017**, *17*, 2260-2270.
6. Perrior, T. Overcoming bottlenecks in drug discovery. *Drug Discovery World* **2010**, 29-33.
7. Schneider, G. Automating drug discovery. *Nat Rev Drug Discov* **2018**, *17*, 97-113.
8. Bleicher, K. H.; Böhm, H. J.; Müller, K.; Alanine, A. I. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2003**, *2*, 369-378.
9. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational methods in drug discovery. *Pharmacol Rev* **2014**, *66*, 334-395.

10. Wang, T.; Wu, M. B.; Lin, J. P.; Yang, L. R. Quantitative structure-activity relationship: promising advances in drug discovery platforms. *Expert Opin Drug Discov* **2015**, *10*, 1283-1300.
11. Guha, R. On exploring structure-activity relationships. *Methods Mol Biol* **2013**, *993*, 81-94.
12. Ślędź, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* **2018**, *48*, 93-102.
13. Bienstock, R. J. Computational methods for fragment-based ligand design: growing and linking. *Methods Mol Biol* **2015**, *1289*, 119-135.
14. Kalyaanamoorthy, S.; Chen, Y. P. Structure-based drug design to augment hit discovery. *Drug Discov Today* **2011**, *16*, 831-839.
15. Hartenfeller, M.; Schneider, G. De novo drug design. *Methods Mol Biol* **2011**, *672*, 299-323.
16. Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K. H.; Schneider, G.; Jacoby, E.; Renner, S. A collection of robust organic synthesis reactions for *in silico* molecule design. *J Chem Inf Model* **2011**, *51*, 3093-3098.
17. Spring, D. R. Diversity-oriented synthesis; a challenge for synthetic chemists. *Org Biomol Chem* **2003**, *1*, 3867-3870.
18. Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964-1969.
19. Tan, D. S. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat Chem Biol* **2005**, *1*, 74-84.
20. O' Connor, C. J.; Beckmann, H. S.; Spring, D. R. Diversity-oriented synthesis: producing chemical tools for dissecting biology. *Chem Soc Rev* **2012**, *41*, 4444-4456.
21. Collins, I.; Jones, A. M. Diversity-oriented synthetic strategies applied to cancer chemical biology and drug discovery. *Molecules* **2014**, *19*, 17221-17255.
22. Lee, J. S.; Lee, J. W.; Kang, N.; Ha, H. H.; Chang, Y. T. Diversity-oriented approach for chemical biology. *Chem Rec* **2015**, *15*, 495-510.
23. Kim, J.; Jung, J.; Koo, J.; Cho, W.; Lee, W. S.; Kim, C.; Park, W.; Park, S. B. Diversity-oriented synthetic strategy for developing a chemical modulator of protein-protein interaction. *Nat Commun.* **2016**, *7*, 13196.
24. Wang, Y.; Wach, J. Y.; Sheehan, P.; Zhong, C.; Zhan, C.; Harris, R.; Almo, S. C.; Bishop, J.; Haggarty, S. J.; Ramek, A.; Berry, K. N.; O'Herin, C.; Koehler, A. N.; Hung, A. W.; Young, D. W. Diversity-oriented synthesis as a strategy for fragment evolution against GSK3 $\beta$ . *ACS Med Chem Lett* **2016**, *7*, 852-856.
25. Gerry, C. J.; Schreiber, S. L. Chemical probes and drug leads from advances in synthetic planning and methodology. *Nat Rev Drug Discov* **2018**, *17*, 333-352.
26. Hoffer, L.; Renaud, J. P.; Horvath, D. In silico fragment-based drug discovery: setup and validation of a fragment-to-lead computational protocol using S4MPLE. *J Chem Inf Model* **2013**, *53*, 836-851.
27. Raux, B.; Voitovich, Y.; Derviaux, C.; Lugari, A.; Rebuffet, E.; Milhas, S.; Priet, S.; Roux, T.; Trinquet, E.; Guillemot, J. C.; Knapp, S.; Brunel, J. M.; Fedorov, A. Y.; Collette, Y.; Roche, P.; Betzi, S.; Combes, S.; Morelli, X. Exploring selective inhibition of the first bromodomain of the human bromodomain and extra-terminal domain (BET) proteins. *J Med Chem* **2016**, *59*, 1634-1641.
28. Kozakov, D.; Hall, D. R.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; Vajda, S. Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc Natl Acad Sci U S A* **2015**, *112*, E2585-2594.
29. Nitsche, C.; Otting, G. NMR studies of ligand binding. *Curr Opin Struct Biol* **2018**, *48*, 16-22.

- 1  
2  
3 30. Merino, F.; Raunser, S. Electron cryo-microscopy as a tool for structure-based drug development. *Angew Chem Int Ed Engl* **2017**, *56*, 2846-2860.
- 4  
5 31. Chevillard, F.; Kolb, P. SCUBIDOO: a large yet screenable and easily searchable database of  
6 computationally created chemical compounds optimized toward high likelihood of synthetic  
7 tractability. *J Chem Inf Model* **2015**, *55*, 1824-1835.
- 8  
9 32. Batiste, L.; Unzue, A.; Dolbois, A.; Hassler, F.; Wang, X.; Deerain, N.; Zhu, J.; Spiliotopoulos, D.;  
10 Nevado, C.; Caflich, A. Chemical space expansion of bromodomain ligands guided by in silico virtual  
11 couplings (AutoCouple). *ACS Cent Sci* **2018**, *4*, 180-188.
- 12  
13 33. Hoffer, L.; Chira, C.; Marcou, G.; Varnek, A.; Horvath, D. S4MPLE-sampler for multiple protein-  
14 ligand entities: methodology and rigid-site docking benchmarking. *Molecules* **2015**, *20*, 8997-9028.
- 15  
16 34. Yang, L.; Tan, C. H.; Hsieh, M. J.; Wang, J.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R.  
17 New-generation amber united-atom force field. *J Phys Chem B* **2006**, *110*, 13166-13176.
- 18  
19 35. Mukherjee, G.; Patra, N.; Barua, P.; Jayaram, B. A fast empirical GAFF compatible partial atomic  
20 charge assignment scheme for modeling interactions of small molecules with biomolecular targets. *J*  
21 *Comput Chem* **2011**, *32*, 893-907.
- 22  
23 36. Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M.; Lehmann, J. W.;  
24 Morehouse, G. F.; Burke, M. D. Synthesis of many different types of organic small molecules using  
25 one automated process. *Science* **2015**, *347*, 1221-1226.
- 26  
27 37. Dittrich, P. S.; Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nat Rev Drug Discov* **2006**,  
28 *5*, 210-218.
- 29  
30 38. Liu, Z.; Wang, P.; Chen, H.; Wold, E. A.; Tian, B.; Brasier, A. R.; Zhou, J. Drug discovery targeting  
31 bromodomain-containing protein 4. *J Med Chem* **2017**, *60*, 4533-4558.
- 32  
33 39. Stathis, A.; Bertoni, F. BET proteins as targets for anticancer treatment. *Cancer Discov* **2018**, *8*, 24-  
34 36.
- 35  
36 40. Abad-Zapatero, C.; Perišić, O.; Wass, J.; Bento, A. P.; Overington, J.; Al-Lazikani, B.; Johnson, M. E.  
37 Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an  
38 atlas-like representation. *Drug Discov Today* **2010**, *15*, 804-811.
- 39  
40 41. Degorce, F.; Card, A.; Soh, S.; Trinquet, E.; Knapik, G. P.; Xie, B. HTRF: A technology tailored for  
41 drug discovery - a review of theoretical aspects and recent applications. *Curr Chem Genomics* **2009**,  
42 *3*, 22-32.
- 43  
44 42. Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The role of ligand  
45 efficiency metrics in drug discovery. *Nat Rev Drug Discov* **2014**, *13*, 105-121.
- 46  
47 43. Tarcsay, A.; Nyíri, K.; Keserü, G. M. Impact of lipophilic efficiency on compound quality. *J Med*  
48 *Chem* **2012**, *55*, 1252-1260.
- 49  
50 44. Shultz, M. D. Setting expectations in molecular optimizations: strengths and limitations of  
51 commonly used composite parameters. *Bioorg Med Chem Lett* **2013**, *23*, 5980-5891.
- 52  
53 45. Shultz, M. D. The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic  
54 optimizations. *Bioorg Med Chem Lett* **2013**, *23*, 5992-6000.
- 55  
56 46. Scott, J. S.; Waring, M. J. Practical application of ligand efficiency metrics in lead optimisation.  
57 *Bioorg Med Chem* **2018**. In press.
- 58  
59 47. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in  
60 medicinal chemistry. *Nat Rev Drug Discov* **2007**, *6*, 881-890.

- 1  
2  
3 48. Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs3: a web server  
4 for compound property calculation and chemical library design. *Nucleic Acids Res* **2015**, *43*, W200-  
5 207.
- 6 49. Hoffer, L.; Horvath, D. S4MPLE-sampler for multiple protein-ligand entities: simultaneous docking  
7 of several entities. *J Chem Inf Model* **2013**, *53*, 88-102.
- 8  
9 50. Filippakopoulos, P.; Knapp, S. The bromodomain interaction module. *FEBS Lett* **2012**, *586*, 2692-  
10 2704.
- 11 51. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D*  
12 *Biol Crystallogr* **2010**, *66*, 133-144.
- 13 52. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.
- 14 53. Collaborative Computational Project, N. m. The CCP4 suite: programs for protein crystallography.  
15 *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 760-763.
- 16 54. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta*  
17 *Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501.
- 18 55. Nicholls, R. A. Ligand fitting with CCP4. *Acta Crystallogr D Struct Biol* **2017**, *73*, 158-170.
- 19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1**

Compound <sup>a</sup>	VS Rank <sup>b</sup> (top %)	pIC <sub>50</sub> <sup>d</sup>	Log improvement
F1	-	4.9±0.02	0.0
F2	-	4.6±0.05	-0.3
1	6 (1.0)	6.4±0.02	1.5
2	10 <sup>c</sup> (1.7)	6.2±0.01	1.3
3	10 <sup>c</sup> (1.7)	6.2±0.03	1.3
4	30 (5.2)	6.0±0.02	1.1
5	40 <sup>c</sup> (6.9)	6.3±0.02	1.4
6	40 <sup>c</sup> (6.9)	6.1±0.03	1.2
7	50 (8.7)	6.5±0.07	1.6
8	52 (9.0)	6.1±0.03	1.2
9	59 <sup>c</sup> (10.2)	6.4±0.02	1.5
10	59 <sup>c</sup> (10.2)	6.3±0.02	1.4
11	65 (11.3)	5.8±0.07	0.9
12	69 (12.0)	5.7±0.05	0.8
13	71 (12.3)	6.2±0.02	1.3
14	75 (13.0)	6.3±0.06	1.4
15	75 (13.0)	5.3±0.03	0.4
16	80 (13.9)	5.9±0.05	1.0
17	89 (15.5)	6.6±0.02	1.7

<sup>a</sup> Compounds discussed in the text, see **Fig. S2** for 2D structures of these compounds.

<sup>b</sup> Absolute ranking position after virtual screening and relative position in parenthesis.

<sup>c</sup> Several compounds show the same interaction energy value and are given the same rank.

<sup>d</sup> pIC<sub>50</sub> values measured by HTRF in triplicates.

**Table 1.** *In silico* ranking and HTRF evaluation data for the 17 compounds selected through virtual screening. **F1** corresponds to the initial hit fragment and **F2** to the activated starting core for the hit-to-lead improvement.

**Table 2**

Compound	$\Delta T_m$ ( $^{\circ}\text{C}$ ) <sup>a</sup>	$K_D$ ( $\mu\text{M}$ ) <sup>b</sup>
1	1.50 $\pm$ 0.00	na <sup>c</sup>
2	3.50 $\pm$ 0.00	0.59 $\pm$ 0.30
3	3.00 $\pm$ 0.00	0.56 $\pm$ 0.13
4	3.00 $\pm$ 0.00	0.94 $\pm$ 0.05
5	2.88 $\pm$ 0.25	0.32 $\pm$ 0.09
6	2.63 $\pm$ 0.25	0.54 $\pm$ 0.09
7	3.13 $\pm$ 0.25	0.36 $\pm$ 0.09
8	2.13 $\pm$ 0.25	0.86 $\pm$ 0.27
9	3.00 $\pm$ 0.00	0.31 $\pm$ 0.05
10	3.38 $\pm$ 0.25	0.41 $\pm$ 0.09
13	3.00 $\pm$ 0.00	0.77 $\pm$ 0.12
14	3.25 $\pm$ 0.29	0.89 $\pm$ 0.16
17	3.88 $\pm$ 0.25	0.19 $\pm$ 0.07

<sup>a</sup>  $\Delta T_m$  values were measured using a thermal shift assay in triplicates.

<sup>b</sup>  $K_D$  values were measured by ITC in triplicates.

<sup>c</sup> no reproducible  $K_D$  value could be measured for this compound.

**Table 2. Orthogonal validation of the most improved compounds.** Compounds exhibiting affinity values below 1  $\mu\text{M}$  by HTRF were purified and further evaluated using orthogonal validation with thermal shift assay (TSA) and isothermal titration calorimetry (ITC).

## Figure legends

**Figure 1. Schematic workflow of the DOTS pipeline.** **1.** The binding mode of a hit molecule is identified by X-ray crystallography. **2.** One or more chemical reactions compatible with the hit are selected from a list of medicinal chemistry-relevant reactions. **3.** A virtual diversity-oriented chemical library is designed using the selected reaction(s) and a collection of compatible commercially available BBs. **4.** The chemical database is docked into the target structure using constrained docking with the S4MPLE conformational sampling tool based on the 3D structure of the target-hit complex. **5.** The poses are ranked using a force-field-based energy function. **6.** Compounds are selected in the top list, and corresponding BBs are purchased. **7.** The desired compounds are synthesized in parallel using an automated robotic platform. **8.** The newly synthesized products are transferred to a Labcyte Access/Echo<sup>®</sup> Laboratory Workstation to be evaluated using homogeneous time-resolved fluorescence (HTRF<sup>®</sup>) technology, isothermal titration calorimetry (ITC) or a thermal shift assay (TSA). Experimental and *in silico* steps are depicted in light orange and light blue, respectively.

**Figure 2. *In silico* workflow of the DOTS pipeline.** Focused chemical library design and VS protocol **A. Focused chemical library design.** The raw focused library is automatically computer-generated by coupling an activated form of the initial fragment with a collection of appropriate BBs selected from a preprocessed database (Fig. S7). The coupling scheme is selected from a set of robust SMARTS-encoded organic synthesis reactions selected by medicinal chemists. This computation stage uses routines from the RDKit package, and the output is a 1D-SMILES file. **B. Post-processing of the raw focused library.** The goal of this stage is to convert 1D-SMILES compounds into 3D-SDF compounds that can be screened with the S4MPLE tool. Compounds are first submitted to common standardization and structural checks (purple box), resulting in a 2D-SDF file. Several filtering steps are then applied to generate a library compliant with medicinal chemistry rules (red boxes). Compounds containing predefined privileged scaffolds can be selected (green box). For large chemical libraries, a clustering stage can be applied to reduce the number of compounds and

1  
2  
3 generate a chemically diverse library (orange box). These latter stages, represented by dotted boxes,  
4  
5 are optional. Several computational stages are then performed to generate the 3D-SDF library (blue  
6  
7 boxes), with all compounds superimposed onto the reference fragment. Finally, atoms that are  
8  
9 restrained during the VS stage are flagged, and GAFF atom types are attributed (cyan boxes). All of  
10  
11 these steps are fully automated. **C. VS of the focused library.** Each compound is prepositioned within  
12  
13 the active site by superposition with the initial fragment, and atomic restraints are applied during  
14  
15 sampling with S4MPLE to preserve the original binding mode. All constraints are eventually removed  
16  
17 before energy minimization of all nonredundant poses after the sampling is completed. The  
18  
19 conformational sampling is also performed for each compound without the protein to identify its  
20  
21 lowest energy conformer in the free state. The ranking for each compound is then achieved by  
22  
23 computing the difference between the best bound and free potential energies. Compounds  
24  
25 displaying a RMSD shift after energy minimization larger than a user-defined threshold are discarded.

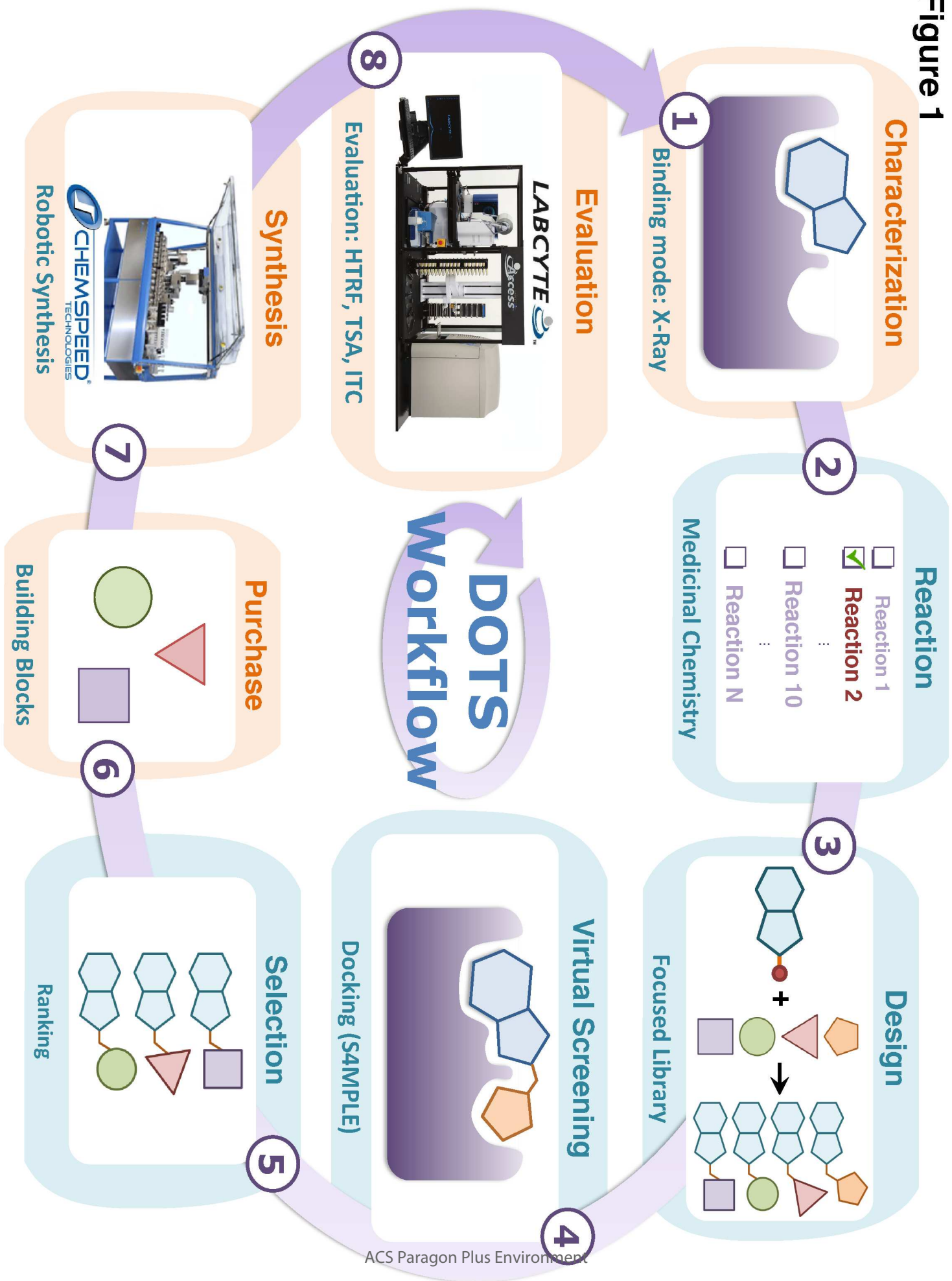
26  
27  
28 **Figure 3. Experimental evaluation of compound 17.** **A.** 2D structure of compound **17** (4-[(2-oxo-  
29  
30 2,3,4,5-tetrahydro-1H-1-benzazepine-7-sulfonamido)methyl]-3-benzyl-7-ethyl-1H-purine-2,6(3H,7H)-  
31  
32 dione)). **B.** Homogeneous time-resolved fluorescence (HTRF) of **17** on BRD4(BD1) after excitation at  
33  
34 337 nm, energy transfer at 620 nm, and fluorescence emission at 665 nm. Fluorescence data are  
35  
36 normalized and plotted as a function of the ligand concentration. The IC<sub>50</sub> value is indicated in  
37  
38 nanomolar. **C.** Isothermal titration calorimetry (ITC) thermogram (top) and nonlinear least-squares fit  
39  
40 model of the integrated data (bottom) at 15°C (100 μM BRD4(BD1) vs. 40 μM compound **17**) as a  
41  
42 function of the molar ratio of the ligand (cell) to the protein (syringe). The K<sub>D</sub> value is indicated in  
43  
44 nanomolar.  
45  
46

47  
48 **Figure 4. Comparison of the binding mode of F1 and compound 17.** **A.** 2D structure of fragment **F1**  
49  
50 (3-benzyl-7-ethyl-1H-purine-2,6(3H,7H)-dione). **B.** 2D structure of **17** (4-[(2-oxo-2,3,4,5-tetrahydro-  
51  
52 1H-1-benzazepine-7-sulfonamido)methyl]-3-benzyl-7-ethyl-1H-purine-2,6(3H,7H)-dione)). **C.**  
53  
54 Superimposition of **F1** (purple) and **17** (orange) in the N-acetylated lysine pocket. **D-G.** 3D  
55  
56 crystallographic structures of **F1** (purple) and **17** (orange) in complex with BRD4(BD1) obtained at 1.2  
57  
58  
59  
60

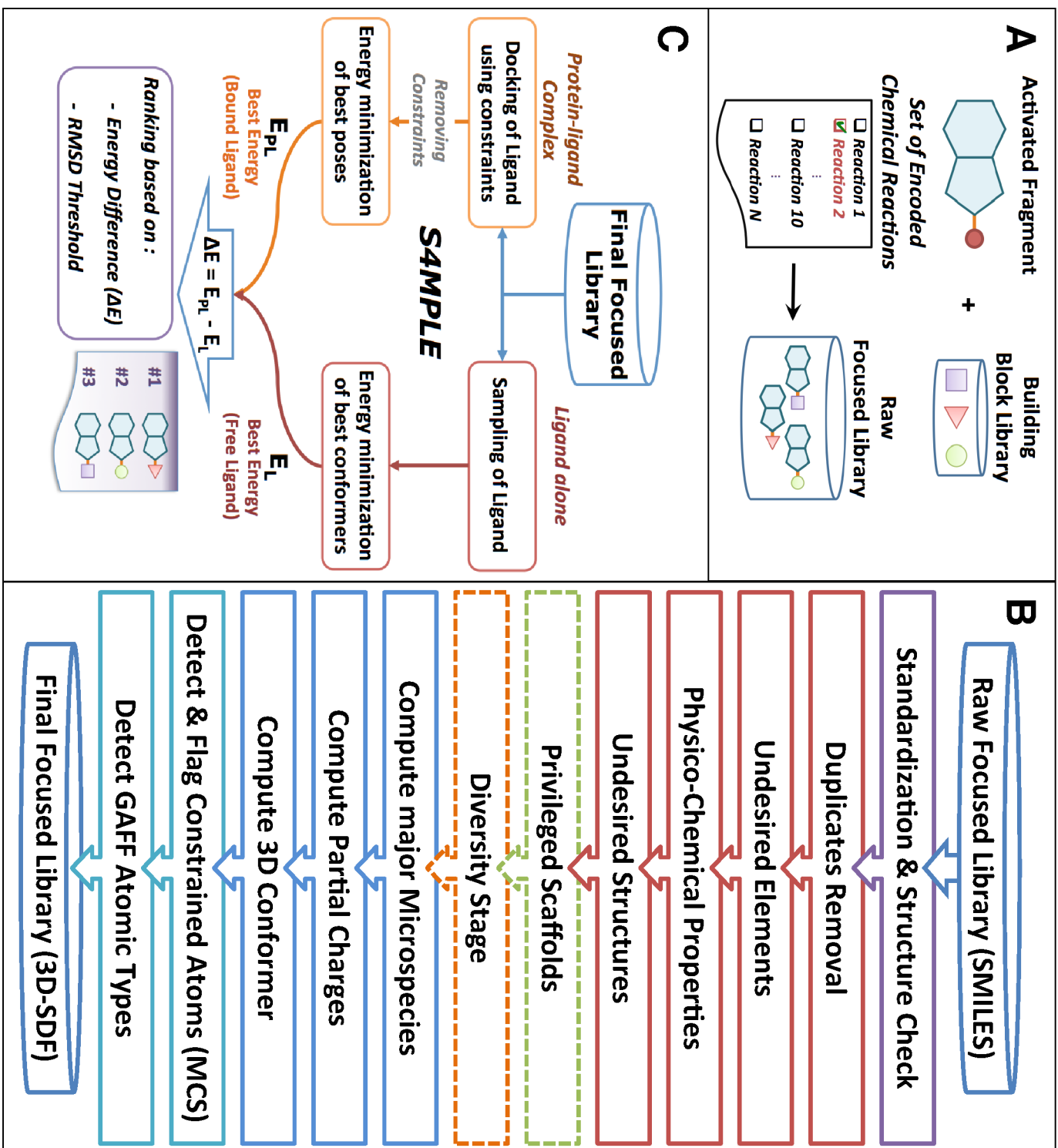
1  
2  
3 Å (PDB code: 6FNX) and 0.95 Å (PDB code: 6FO5) resolution, respectively. Ligands are displayed in  
4 ball-and-stick representation. Key residues of BRD4(BD1) are represented as sticks and colored green  
5 for the WPF shelf (residues 81-83), orange for Y97, pink for N140, blue for the gatekeeper I146 and  
6 white for the others. Water molecules are represented by red spheres. Direct hydrogen bonds  
7 between the ligand and protein are represented by gray dashed lines. **D-E.** The network among water  
8 molecules located in the cavity is represented by red dashed lines. The indirect bonds between ligand  
9 and protein mediated by water molecules are represented by orange dashed lines. **F-G.** Van der  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

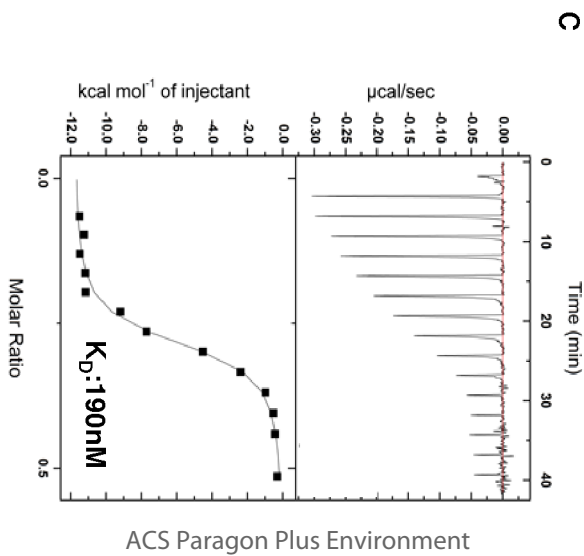
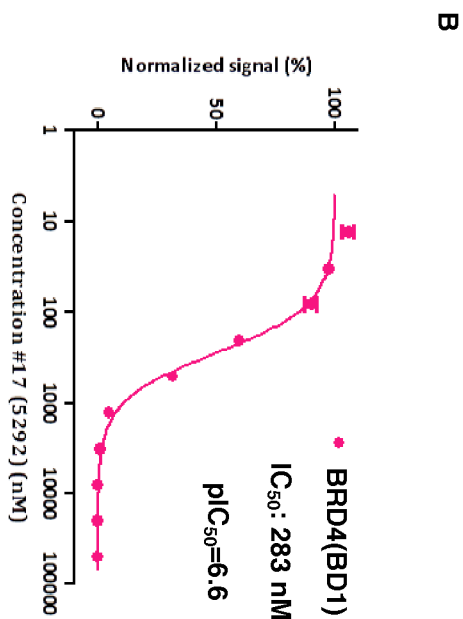
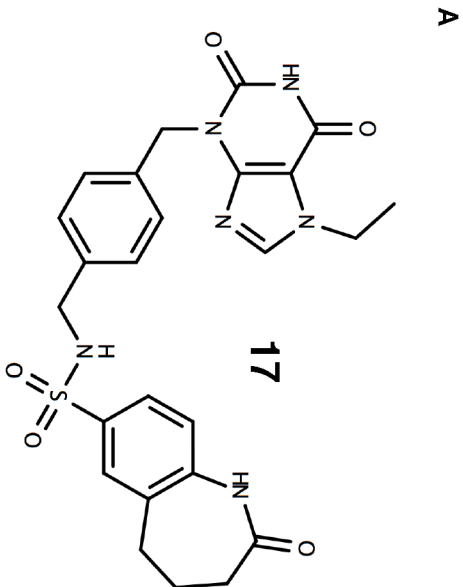
Waals contacts between the ligand and protein are represented by light-blue dashed lines.

Figure 1



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

