CovaDOTS: In Silico Chemistry-Driven Tool to Design Covalent Inhibitors Using a Linking Strategy

Laurent Hoffer, Magali Saez-Ayala, Dragos Horvath, Alexandre Varnek, Xavier Morelli, Philippe Roche

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

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

[a]CRCM, CNRS, INSERM, Institut Paoli-Calmettes, Aix-Marseille Univ, Marseille, France

[b]Department of Organic Chemistry, Lobachevsky State University of Nizhni Novgorod, 23 Gagarin Avenue, Nizhni Novgorod, 603950, Russia

[c]Laboratoire de Chemoinformatique, CNRS UMR7140, 1 rue Blaise Pascal, 67000, Strasbourg, France

[d]IPC Drug Discovery Platform, Institut Paoli-Calmettes, 232 Boulevard de Sainte-Marguerite, 13009, Marseille, France

[1]Laurent Hoffer and Yuliia Voitovich contributed equally to this work

[*]To whom correspondence should be addressed. E-mail: philippe.roche@inserm.fr 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.[1, 2] 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.[3] 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.[4, 5]

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.[6, 7]

This optimization phase to improve the affinity, selectivity and physicochemical properties of the compounds involves a combination of experimental and *in silico* approaches.[8, 9] 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

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

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

Synthetic approaches, such as diversity-oriented synthesis (DOS), have been developed to generate structural diversity in a rational and efficient manner. The goal of such strategies is to maximize the chemical space explored while limiting the number of compounds to a minimum.[17, 18] DOS concept has been applied with success for the discovery of biologically active compounds and their evolution as probes.[19-25]

In this study, we report a H2L optimization strategy integrating computational and automated experimental methods: DOTS (Diversity-Oriented Target-focused Synthesis). A general pipeline has been implemented to design more potent chemical probes starting from an initial fragment. First, the binding mode of the initial hit is identified using X-ray crystallography. Then, the *in silico*

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

DOTS was successfully applied to optimize a recently characterized inhibitor of bromodomain (BRD), a protein from the bromodomain and extra-terminal (BET) family.[27] Several new compounds exhibiting improved affinities (IC_{50} and/or K_D values) have been synthesized and validated starting from the core fragment. The X-ray structure of the most potent compound was solved and used to decipher the molecular basis of the improved affinity.

Results and Discussion

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

The minimum common substructure necessary to bind the target is identified using either fragment-based approaches or deconstruction of an identified hit compound.[5, 28] The binding mode is characterized by structural models of the fragment complexed with the biological target using X-ray crystallography (Fig. 1.*1*). Other biophysical methods could also be used at this stage, such as NMR or cryo-EM.[29, 30] The common core is used to build a virtual chemical library focused on the hit compound (Figs. 1.*2* and 1.*3*). The design of this diversity-oriented chemical library is achieved by

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

Building virtual focused libraries

Progresses in computational methods have made it possible to screen millions of virtual compounds *in silico* in a reasonable amount of time. However, computer-generated compounds are not always commercially available or easily accessible to organic synthesis. It is therefore crucial to incorporate well-recognized medicinal chemistry rules in the design of virtual libraries.[31, 32]

DOTS uses a relevant set of robust organic synthesis reactions defined by Hartenfeller and coworkers or implemented in house.[16] These rules, designed by both medicinal chemists and molecular modelers, include the most common and well-accepted reactions encountered during the H2L stages in the pharma industry. Most of these rules involve only a one-step reaction, and half of them concern cycle formation. The first step of virtual library design is the careful choice of one or a few chemical reactions compatible with the starting material from the list of available chemical reactions (Figs. 1.*2* and 2.*A*). Consequently, the initial hit must be activated to include a reactive functionality that matches at least one reaction in the set of medicinal chemistry-relevant reactions. A collection of functionalized BBs compatible with the selected reaction(s) is assembled from commercially available databases. The activated hit is coupled *in silico* to the different BBs to generate the focused chemical library (Figs. 1.*3* and 2.*A*). This strategy almost guarantees that compounds generated in the virtual library are easily amenable to organic synthesis in one or two steps with high expected yields and no byproducts. In addition, all compounds in the virtual library contain the original hit scaffold as a submoiety, making it a focused library. Several fully automated post-processing stages are applied to the virtual library to extract a diverse subset of duplicate-free representative compounds with a focus on medicinal chemistry-relevant structures with reasonable physiochemical properties (Fig. 2.*B,* see supporting information).

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.[33] S4MPLE uses the AMBER energy force field[34] and GAFF[35] 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.[13] 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.[36, 37] 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

that are predicted to bind to the biological target are synthesized as opposed to standard combinatorial chemistry approaches.

Optimization of BRD inhibitors with DOTS: proof of concept

BRDs are protein domains that specifically recognize acetylated lysine (Kac) residues, such as those present on the histone tails. The BRD4 protein contains two BRD in tandem and belongs to the bromodomain and extra-terminal (BET) family.[38] These epigenetic readers of lysine acetylation emerged as druggable targets due to their role in inflammatory and cardiovascular diseases and cancer.[39] Thus, developing molecules that could efficiently disrupt this protein-protein interaction is highly desirable.

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

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

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

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

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

Selection of compounds to be synthesized: The prioritization of the putative optimized compounds was based on several criteria, the most important of which was the force-field energy difference between the bound and free states of the compound. Only the top 100 compounds were considered for further analysis. Among those, compounds that did not converge towards the targeted pocket area (F79, D145, L148, M149 and A152) and compounds exhibiting a shift from the initial binding mode were not considered. We performed a clustering of compounds based on structural similarities and selected only one representative compound for each cluster (see supplementary information). Remaining poses were subjected to visual inspection. When no additional strong interaction was detected, the less substituted compound was selected for the synthesis stage. At the end of this process, 17 representative molecules were selected for the subsequent synthesis stage (Fig. S2 and Table 1). Each of these compounds did not necessarily correspond to the best-ranked one within a given cluster but was predicted to have the highest ligand efficiency value.[40]

Robotic synthesis: The 17 compounds were generated using a sulfonylation reaction of benzylamino-containing xanthine derivative **F2** with appropriate sulfonyl chloride BBs (Fig. S1). All compounds

have been efficiently prepared in high yields (87 ± 11%) using our automated robotic platform from Chemspeed.

In vitro evaluation of the compounds

The crude synthetic products were directly transferred to a Labcyte Access/Echo® Laboratory Workstation to assess the compounds for their ability to disrupt BRD/histone complexes using a homogeneous time-resolved fluorescence (HTRF)-based assay.[41] The primary screening results revealed that, compared with the original hit, all 17 *de novo*-designed molecules displayed improved pIC_{50} (Table 1). The initial hit (**F1**) exhibited a pIC_{50} value of 4.9, whereas the amino activated fragment (**F2**) demonstrated a pIC_{50} value of 4.6. Fourteen compounds (82%) exhibited an affinity improvement greater than 1 log, which is the expected range of improvement in a classical round of hit optimization. They all contained an aromatic ring directly attached to the sulfonyl group. Among the most improved compounds in terms of affinity were biphenyl derivative (**1**) and 4-benzyloxyphenyl-sulfonamide derivative (**7**) with pIC_{50} values 6.4 and 6.5, respectively. Tetrahydroquinolinone derivative **9** exhibited high inhibitory activity (pIC_{50} = 6.4) whereas its benzazepin-2-one analog (**17**) exhibited the best target affinity with a pIC_{50} value of 6.6, an approximately 2-log improvement over the initial hit **F1**.

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

Orthogonal validation of the compounds

To further validate the approach, all compounds exhibiting submicromolar affinities in HTRF were resynthesized using a state-of-the-art approach, including a purification step, and analyzed using ITC and TSA as orthogonal validation methods (Table 2). All thirteen purified compounds were confirmed as high affinity binders by both methods (Fig. S3). Submicromolar K_D values consistent with the HTRF data were measured for all but one compound, and the best compound (**17**) showed an affinity of

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

Beside optimization of the ligand affinity for the target, it is important to take into account ADME parameters during the H2L process. During the preparation of the virtual library several filters were applied using molecular descriptors that are important to predict the drug-likeness of compounds. This increases the probability that the designed compounds possess physicochemical properties that are compatible with further developments. Besides, different efficiency metrics have been proposed to estimate the potency of compounds during the H2L process.[42] Among them, the ligand efficiency (LE), defined as the binding free energy per heavy atom count (LE = ΔG/HAC). All 13 validated compounds show decreased LE values compared to the initial fragment (Table S1). It is not uncommon in fragment to lead optimization programs to observe such drops in LE values.[43, 44] The best LE (0.27 kcal/mole/heavy atom) corresponds to 4-methyl-benzene sulfonamide derivative (13), the less substituted compound. However, all compounds exhibit similar LE values (0.24 ± 0.01 kcal/mole/heavy atom) and therefore this metrics cannot be used to prioritize compounds. The lipophilic ligand efficiency (LLE) is an estimate of the contribution of lipophilicity to potency (LLE = pIC_{50} - cLogP). LLE has been shown to improve during fragment to lead optimization programs and suggested as a valid metrics to prioritize compounds.[42-47]. More than half of the validated compounds demonstrate better LLE compared to the initial hit **F1** (Table S1) with compounds **8**, **9** and **17** being the most improved. Finally, the aqueous solubility of the validated compounds has been measured (Table S1). The most soluble compounds are **17**, **8**, **9**, **13** and **3**. Taking into account these different results, **17** appears the most appealing compound for further developments although several other compounds show interesting potential.

The X-ray structure of the most improved compound in terms of binding affinity (**17**) complexed with the BRD4(BD1) protein target was solved (PDB code: 6FO5) to confirm the binding mode predicted by the modeling experiments and to identify the new interactions responsible for the improvement in affinity. The final 2Fo − Fc omit map showed clear and unambiguous density throughout the ligand

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

Conclusion

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

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

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

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

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

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

The synergistic approach of DOTS requires an optimal combination of expertise from modelers, chemists and biologists to better explore the chemical space around given hits. Human medicinal and organic chemistry expertise is needed to select relevant chemical reactions, whereas computational

methods make it possible to limit the number of compounds to be synthesized and evaluated only to those that are predicted to bind to the target efficiently.

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

Experimental Section.

Preparation of building blocks

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

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

Chemical reactions encoding

The 58 chemical reactions described by Hartenfeller encoded in SMARTS format were retrieved.[16] Most of these rules are designed as "Reactant1 + Reactant2 → Product" and half of them correspond to cycle formation. For the bromodomain test case, reaction corresponding to the coupling of a sulfonylchloride and free amino groups to form a sulfonamide was selected (Fig. S1.*A*). In house reactions have also been implemented to handle alternative synthesis schemes in order to design putative suicide inhibitors involving a cysteine residue.

Focused chemical library design

An in house python script, based on the RDKit toolkit (http://www.rdkit.org) was used to create the raw focused library in SMILES format. BBs containing a sulfonyl-chloride group were automatically retrieved from the pre-processed MolPort database. Compound **F2** containing the primary amine reactive group, was coupled to the ensemble of 1200 sulfonylchloride-containing BBs to generate the virtual library (Fig. S1.*B*). The post-processing of this raw focused library is described in the Fig. 2.*B*. A structural check of generated compounds was performed with StructureChecker to eliminate compounds with warnings. Putative duplicates and compounds that contain at least one undesired element (other than C, H, N, O, P, S, F, Cl, Br or I) were also rejected. The next filtering step is based on physico-chemical properties in order to discard structures with obvious warnings. Common threshold values were used (cLogP ≤5, tPSA ≤150 Å², rotational bonds ≤10 and ring system size ≤2). Due to both the size of original **F2** fragment and the limited number of sulfonylchloride-containing BBs, a greater value for the maximal molecular weight was employed (575 Da by contrast to the default 500 Da value). Undesired compounds were excluded based on yuck filters used to generate

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

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

Conformational sampling

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

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

Compound selection

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

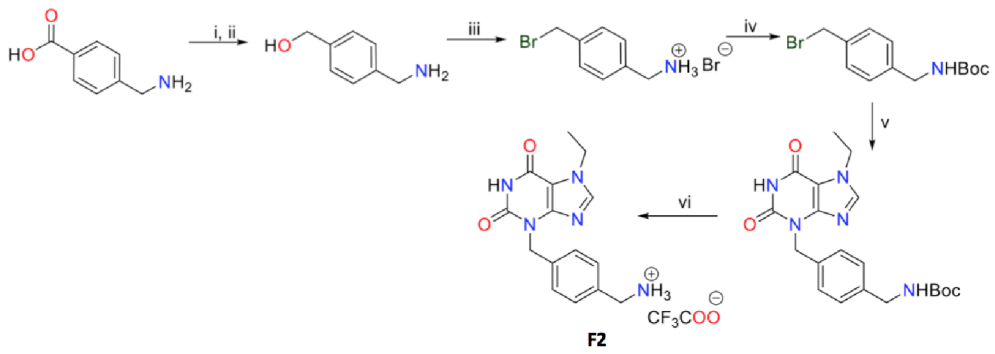
Chemical synthesis

Commercially available reagents were used without additional purification. Automated syntheses were performed by using a Robotic platform Accelerator Synthesizer SLTII from Chemspeed technologies. Column chromatography was performed using Macherey-Nagel Kieselgel 60 (70-230 mesh). The petroleum spirit refers to the fraction with distillation range 40-70°C. ^{1}H and ^{13}C NMR

spectra were recorded at room temperature in DMSO-*d6*, CD_3OD or $CDCl_3$ by using a Bruker AC400, AC250 or Agilent DD2 400 spectrometers. Chemical shifts (δ) are reported in parts per million (ppm) with internal reference TMS and coupling values (J) in hertz. Abbreviations for peaks are, br: broad, s: singlet, d: doublet, t: triplet, q: quadruplet, quint: quintuplet, sex: sextuplet and m: multiplet. The spectra recorded are consistent with the proposed structures. Reaction monitoring and purity of compounds were recorded by using analytical Agilent Infinity high performance liquid chromatography (Column Zorbax SB-C18 1.8 µM (2.1x50 mm); Mobile phase (A: 0.1% FA H_2O, B: 0.1% FA MeCN, Time/%B 0/10, 4/90, 7/90, 9/10, 10/10); Flow rate 0.3 mL/min; Diluent MeOH) with DAD at 230 nM. All tested compounds yielded data consistent with a purity of ≥ 95%.

Scheme 1

Scheme 1. Synthesis of 3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt F2[a]



[a]reagents and conditions: i) $LiAlH_4$, THF, 0°C-reflux, 12 hours; ii) H_2O, NaOH, 0°C, 1h; iii) HBr, H_2O, reflux, 3.5 hours; iv) Boc_2O, $NaHCO_3$, H_2O-dioxane, rt, 3 hours; v) 7-ethyl-1H-purine-2,6(3H,7H)-dione, K_2CO_3, rt, 2 days; vi) CF_3COOH, CH_2Cl_2, rt, 1 hour.

(4-(aminomethyl)phenyl)methanol. To a stirred suspension of commercially available 4-(aminomethyl)benzoic acid (10g, 66.2mmol) in tetrahydrofurane (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_2O were added, then 15mL of 10% NaOH and 30mL of H_2O. 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_2O to afford 4-(aminomethyl)benzyl alcohol **bb** (8.8g, yield 97%) as a light-beige powder. ^1H NMR (250 MHz, CD_3OD) δ 7.32 (s, 4H), 4.58 (s, 2H), 3.78 (s, 2H). ^{13}C NMR (63 MHz, CD_3OD) 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_2O (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,

and the solid obtained was washed with Et_2O to afford (4-(bromomethyl)phenyl)methanamine hydrobromic salt as a grey powder (1g, 89%). ^{1}H NMR (250 MHz, DMSO-*d6*) δ 8.26 (brs, 2H), 7.48 (s, 4H), 4.71 (d, *J* = 3.6 Hz, 2H), 4.03 (s, 2H). ^{13}C NMR (63 MHz, DMSO-*d6*) δ 138.5, 134.1, 129.6, 129.4, 41.9, 34.0.

***tert*-butyl 4-(bromomethyl)benzylcarbamate.** To the solution of (4-(bromomethyl)phenyl)methanamine hydrobromic salt (1g, 5mmol) in a mixture of H_2O and dioxane (40mL, 1:1), di-*tert*-butyl dicarbonate (3.3g, 15mmol) was added at 0°C. Then $NaHCO_3$ (0.87g, 10mmol) was added at the same temperature, and the resulting mixture was stirred for 3 hours at room temperature. The mixture was extracted with Et_2O (x3), combined organic layers were dried over Na_2SO_4 and concentrated under reduced pressure. The resulting colorless oil was washed with petroleum ether and the precipitate formed was filtered off and dried under reduced pressure to afford tert-butyl 4-(bromomethyl)benzylcarbamate (1.1g, 73%) as a white solid. ^{1}H NMR (250 MHz, DMSO-*d6*) 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 (s, 9H). ^{13}C NMR (63 MHz, DMSO-*d6*) δ 155.8, 140.5, 136.4, 129.3, 127.2, 77.9, 43.1, 34.6, 28.3.

***tert*-butyl 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate.** To a suspension of 7-ethyl-1H-purine-2,6(3H,7H)-dione (180mg, 1mmol) prepared according to the previously described procedures[27] and K_2CO_3 (151.8mg, 1.1mmol) in dimethylformamide (15mL) *tert*-butyl 4-(bromomethyl)benzylcarbamate (300mg, 1mmol) was added portionwise, and the resulting mixture was stirred at room temperature overnight. The solvent was removed under reduced pressure, and a crude product was purified by column chromatography (eluent 40:1:0.1 CH_2Cl_2:MeOH:NH_3) to afford 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate (202 mg, 51%) as a white solid. ^{1}H NMR (250 MHz, DMSO-*d6*) δ 8.07 (s, 1H), 7.30 (d, *J* = 8.1 Hz, 2H), 7.18 (d, *J* = 8.1 Hz, 2H), 5.09 (s, 2H), 4.25 (q, *J* = 7.1 Hz, 2H), 4.09 (d, *J* = 6.0 Hz, 2H), 1.47-1.31 (m, 12H). ^{13}C NMR (63 MHz, DMSO-*d6*) 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, 16.3.

3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt F2. A suspension of 4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzylcarbamate (200mg, 0.50mmol) in dichloromethane (3mL) was treated with trifluoroacetic acid (450µl), and the resulting mixture was stirred at room temperature for 1 hour. The solvent was removed under reduced pressure, and a crude product was purified by column chromatography (eluent 6:1:0.1, CH_2Cl_2:MeOH:NH_3) to afford 3-(4-(aminomethyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione trifluoroacetic salt **F2** (207mg, 100%) as a white solid. ^{1}H NMR (250 MHz, CD_3OD) δ 7.96 (s, 1H), 7.51 (d, *J* = 8.2 Hz, 2H), 7.42 (d, *J* = 8.2 Hz, 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). ^{13}C NMR (63 MHz, CD_3OD) δ 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.

General procedure for the synthesis of *N*-(4-((7-ethyl-2,6-dioxo-1*H*-purin-3(2*H*,6*H*,7*H*)-yl)methyl)benzyl)sulfonamide derivatives. The automated sulfonylation were run in 2 mL LC glass vials. A stock solution of 3-(4-(aminomethyl)benzyl)-7-ethylxanthine trifluoroacetic salt **F2** (206.5 mg, 1eq) and triethylamine (138.7 µL, 2eq) in anhydrous tetrahydrofuran (20 mL) was prepared and placed on the robotic deck. The robot automatically prepared the reaction mixtures by dispending the prepared solution (400 µL) to the individual vials containing chlorosulfonyl derivatives (0.01 mmol) via the liquid handling tool. After 12 hours at room temperature, samples were taken out and analyzed by LCMS to determine the percent of conversion. The product solutions were concentrated at 80°C, and residues were diluted with dimethylsulfoxide to afford a solution 5.10^{-2} M of the expected compounds. Further LCMS analysis allowed rectifying the final concentration of the products precisely to 10^{-2} M for further biological evaluations.

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

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) LCMS $C_{28}H_{27}N_5O_5S$ Rt = 6.659, m/z found 545.5, calcd 545.6.

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

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

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

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

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

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

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

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

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

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

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

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-3-sulfonamide (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-1*H*-purin-3(2*H*,6*H*,7*H*)-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. **^{1}H NMR** (400 MHz, DMSO) δ 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) δ 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. **^{1}H NMR** (400 MHz, DMSO) δ 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) δ 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.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-(2-methyloxazol-4-yl)benzene-sulfonamide (3) (eluent CH_2Cl_2:MeOH 10:0.5, yield 90%) as white solid. **^{1}H NMR** (250 MHz, DMSO) δ 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 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, 3H), 1.38 (t, *J* = 7.1 Hz, 3H). **^{13}C NMR** (63 MHz, DMSO) δ 162.1, 154.7, 150.8, 149.7, 142.0, 139.4, 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** $C_{25}H_{24}N_6O_5S$, Rt = 6.014, m/z found 521.0, calcd 520.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)benzo[b]thiophene-2-sulfonamide (4) (eluent CH_2Cl_2:MeOH 10:0.5, yield 76%) as light yellow plates. **^{1}H NMR** (400 MHz, DMSO) δ 11.20 (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), 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, 2H), 1.38 (t, *J* = 7.2 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO) δ 154.58, 150.73, 149.63, 141.89, 141.78, 140.63, 137.59, 136.43, 136.07, 128.53, 127.69, 127.59, 127.07, 125.71, 125.42, 122.98, 106.43, 46.04, 44.46, 41.51, 16.17. **LCMS** $C_{23}H_{21}N_5O_4S_2$, Rt = 6.443, m/z found 495.6, calcd 495.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-5-(3-methyl-1H-pyrazol-5-yl)thiophene-2-sulfonamide (5) (eluent CH_2Cl_2:MeOH 10:1, yield 66%) as white solid. **^{1}H NMR** (400 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* = 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, *J* = 7.1 Hz). **^{13}C NMR** (101 MHz, DMSO) δ 154.59, 150.76, 149.66, 141.91, 138.41, 136.60, 136.05, 132.12, 127.70, 127.58, 123.14, 106.45, 101.73, 46.04, 44.51, 41.52, 16.18, 10.48. **LCMS** $C_{23}H_{23}N_7O_4S_2$, Rt = 5.901, m/z found 525.5 calcd 525.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methoxybenzene-sulfonamide (6) (eluent CH_2Cl_2:MeOH 10:0.5, yield 71%) as white solid. **^{1}H NMR** (250 MHz, $CDCl_3$) δ 7.65 – 7.56 (m, 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 (d, *J* = 7.2 Hz, 2H), 3.86 (s, 2H), 3.72 (s, 3H), 1.36 (t, *J* = 7.2 Hz, 3H). **^{13}C NMR** (63 MHz, $CDCl_3$) δ 162.7,

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, 42.4, 16.1. **LCMS** $C_{22}H_{23}N_5O_5S$, Rt = 5.822, m/z found 469.9, calcd 496.5.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-phenoxybenzene-sulfonamide (7) (eluent CH_2Cl_2:MeOH 10:0.5, yield 86%) as white solid. **^{1}H NMR** (250 MHz, $CDCl_3$+CD_3OD) δ 7.71 (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 (s, 2H), 4.26 (q, *J* = 7.1 Hz, 2H), 4.01 (s, 2H), 1.46 (t, *J* = 7.2 Hz, 3H). **^{13}C NMR** (63 MHz, $CDCl_3$+CD_3OD) δ 161.88, 155.66, 155.61, 151.91, 150.70, 141.76, 137.03, 136.02, 134.30, 130.60, 129.56, 128.80, 128.45, 125.35, 120.65, 117.93, 107.94, 46.97, 45.86, 42.91, 16.57. **LCMS** $C_{27}H_{25}N_5O_5S$, Rt = 6.611, m/z found 531.5, calcd 531.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1,3-dimethyl-2-oxo-2,3-dihydro-1H-benzo[d]imidazole-5-sulfonamide (8) (eluent CH_2Cl_2:MeOH 10:0.5, yield 100%) as white solid. **^{1}H 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), 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), 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). **^{13}C NMR** (63 MHz, 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, 107.6, 106.5, 106.1, 55.0, 48.7, 41.6, 27.3, 27.2, 16.3. **LCMS** $C_{24}H_{25}N_7O_5S$, Rt = 5.515, m/z found 524.0, calcd 523.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-2-oxo-1,2,3,4-tetrahydroquinoline-6-sulfonamide (9) (eluent CH_2Cl_2:MeOH 10:1, yield 93%) as white solid. **^{1}H NMR** (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 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* = 7.6 Hz), 2.49 – 2.45 (m), 1.38 (t, *J* = 7.2 Hz). **^{13}C NMR** (101 MHz, DMSO) δ 170.30, 154.57, 150.75, 149.65, 141.90, 141.79, 136.86, 135.90, 133.45, 127.68, 127.46, 126.27, 126.09, 124.05, 114.97, 106.43, 45.82, 44.49, 41.52, 29.85, 24.46, 16.17. **LCMS** $C_{24}H_{24}N_6O_5S$, Rt = 5.353, m/z found 508.5, calcd 508.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-1-phenyl-1H-pyrazole-3-sulfonamide (10) (eluent CH_2Cl_2:MeOH 10:0.7, yield 60%) as light orange solid. **^{1}H NMR** (400 MHz, 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 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).**^{13}C NMR** (101 MHz, DMSO) δ 154.58, 150.75, 149.65, 141.90, 139.13, 138.75, 136.81, 135.97, 129.64, 129.29, 127.75, 127.58, 127.51, 124.86, 119.17, 106.43, 45.88, 44.47, 41.51, 16.17. **LCMS** $C_{24}H_{23}N_7O_4S$, Rt = 6.207, m/z found 505.5, calcd 505.6.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-4-methylbenzene-sulfonamide (13) (eluent CH_2Cl_2:MeOH 10:0.5, yield 79%) as white solid. **^{1}H NMR** (250 MHz, CD_3OD+$CDCl_3$) δ 7.81 (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 (q, *J* = 7.2 Hz, 2H), 3.98 (s, 2H), 2.39 (s, 3H), 1.48 (t, *J* = 7.2 Hz, 3H).**^{13}C NMR** (63 MHz, CD_3OD+$CDCl_3$) δ 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, 43.1, 21.6, 16.7. **LCMS** $C_{22}H_{23}N_5O_4S$, Rt = 5.983, m/z found 453.9, calcd 453.5.

N-(4-((7-ethyl-2,6-dioxo-1H-purin-3(2H,6H,7H)-yl)methyl)benzyl)-[1,1'-biphenyl]-3-sulfonamide (14) (eluent CH_2Cl_2:MeOH 10:0.5, yield 96%) as a white solid. **^{1}H NMR** (400 MHz, DMSO) δ 11.20 (s, 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, 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, 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 Hz, 2H), 1.37 (t, *J* = 7.2 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO) δ 154.57, 150.72, 149.64, 141.87, 141.45, 141.05, 138.67, 136.56, 135.92, 130.40, 129.76, 129.10, 128.18, 127.69, 127.58, 127.52, 126.82, 125.29, 124.46, 106.41, 45.94, 44.46, 41.52, 16.18. **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-oxo-2,3,4,5-tetrahydro-1H-benzo[b]azepine-7-sulfonamide (17) (eluent CH_2Cl_2:MeOH 10:1, yield 69%) as a white solid**. ^{1}H NMR** (250 MHz, DMSO-*d6*) 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,

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), 2.70 (t, J = 6.6 Hz, 2H), 2.24 – 2.01 (m, 4H), 1.37 (t, J = 7.1 Hz, 3H); **^{13}C NMR** (63 MHz, DMSO-*d6*) δ 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), 125.8, 121.6, 106.5, 45.9, 44.5, 41.6, 33.0, 30.0, 27.7, 16.2. **LCMS** $C_{25}H_{26}N_6O_5S$, Rt = 5.370, m/z found 522.5, calcd 522.6.

Robotic platform Labcyte workstation

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

Protein expression and purification

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

For Isothermal Titration Calorimetry (ITC), Thermal Shift Assay (TSA) and crystallogenesis, BRD4(BD1) was produced and purified using a histidine tag affinity chromatography as described by Filipakopoulos *et al.*[50] For these experiments, a pNIC28-BSA4 expression vector containing

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

HTRF assay

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

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

nm)]*$10^{4)}$ with the compound vehicle alone (DMSO) and "min signal" the signal ratio without peptide. For IC_{50} measurements, values were normalized and fitted with Prism 5.03 (GraphPad software) using the following equation: Y = 100 / (1 + ((X / IC_{50})^Hill slope)).

Isothermal titration calorimetry

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

Raw data were scaled after setting the zero to the titration saturation heat value. Integrated raw ITC data were fitted to a one site non-linear least squares fit model using the MicroCal Origin plugin as implemented in Origin 7 (Origin Lab). Finally, ΔG and TΔS values were calculated from the fitted ΔH and K_D values using the equations $\Delta G = -R.T.\ln K_D$ and $\Delta G = \Delta H - T\Delta S$.

Thermal shift assay

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

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

Structural characterization

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

Supporting Information

Additional supplemental tables and figures are provided free of charge on the ACS publication website.

- Generic building of the focused virtual library; 2D structure of all compounds; X-ray structural data for F1 and 17 bound to BRD4(BD1); ITC data of the purified compounds (PDF)

- Molecular formula strings and associated affinity data (CSV)

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 and xavier.morelli@inserm.fr

[1]Laurent Hoffer and Yuliia Voitovich contributed equally to this work

ORCID

laurent.hoffer@inserm.fr, 0000-0003-1906-7128

voitovich_25@mail.ru, 0000-0002-9282-3834

brigitt.raux@gmail.com, 0000-0002-2194-0259

derviauxc@ipc.unicancer.fr, 0000-0001-7441-1126

stephane.betzi@inserm.fr, 0000-0001-5935-5058

dhorvath@unistra.fr, 0000-0003-0173-5714

varnek@unistra.fr, 0000-0003-1886-925X

yves.collette@inserm.fr, 0000-0001-5359-7099

sebastien.combes@univ-amu.fr, 0000-0002-8213-2407

philippe.roche@inserm.fr, 0000-0002-5580-0588

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. Śledź, 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 GSK3β. *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.

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.

31. Chevillard, F.; Kolb, P. SCUBIDOO: a large yet screenable and easily searchable database of computationally created chemical compounds optimized toward high likelihood of synthetic tractability. *J Chem Inf Model* **2015**, *55*, 1824-1835.

32. Batiste, L.; Unzue, A.; Dolbois, A.; Hassler, F.; Wang, X.; Deerain, N.; Zhu, J.; Spiliotopoulos, D.; Nevado, C.; Caflisch, A. Chemical space expansion of bromodomain ligands guided by in silico virtual couplings (AutoCouple). *ACS Cent Sci* **2018**, *4*, 180-188.

33. Hoffer, L.; Chira, C.; Marcou, G.; Varnek, A.; Horvath, D. S4MPLE-sampler for multiple protein-ligand entities: methodology and rigid-site docking benchmarking. *Molecules* **2015**, *20*, 8997-9028.

34. Yang, L.; Tan, C. H.; Hsieh, M. J.; Wang, J.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R. New-generation amber united-atom force field. *J Phys Chem B* **2006**, *110*, 13166-13176.

35. Mukherjee, G.; Patra, N.; Barua, P.; Jayaram, B. A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets. *J Comput Chem* **2011**, *32*, 893-907.

36. Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. Synthesis of many different types of organic small molecules using one automated process. *Science* **2015**, *347*, 1221-1226.

37. Dittrich, P. S.; Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nat Rev Drug Discov* **2006**, *5*, 210-218.

38. Liu, Z.; Wang, P.; Chen, H.; Wold, E. A.; Tian, B.; Brasier, A. R.; Zhou, J. Drug discovery targeting bromodomain-containing protein 4. *J Med Chem* **2017**, *60*, 4533-4558.

39. Stathis, A.; Bertoni, F. BET proteins as targets for anticancer treatment. *Cancer Discov* **2018**, *8*, 24-36.

40. Abad-Zapatero, C.; Perišić, O.; Wass, J.; Bento, A. P.; Overington, J.; Al-Lazikani, B.; Johnson, M. E. Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. *Drug Discov Today* **2010**, *15*, 804-811.

41. Degorce, F.; Card, A.; Soh, S.; Trinquet, E.; Knapik, G. P.; Xie, B. HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr Chem Genomics* **2009**, *3*, 22-32.

42. Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **2014**, *13*, 105-121.

43. Tarcsay, A.; Nyíri, K.; Keseru, G. M. Impact of lipophilic efficiency on compound quality. *J Med Chem* **2012**, *55*, 1252-1260.

44. Shultz, M. D. Setting expectations in molecular optimizations: strengths and limitations of commonly used composite parameters. *Bioorg Med Chem Lett* **2013**, *23*, 5980-5891.

45. Shultz, M. D. The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations. *Bioorg Med Chem Lett* **2013**, *23*, 5992-6000.

46. Scott, J. S.; Waring, M. J. Practical application of ligand efficiency metrics in lead optimisation. *Bioorg Med Chem* **2018**. In press.

47. Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* **2007**, *6*, 881-890.

48. Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res* **2015**, *43*, W200-207.

49. Hoffer, L.; Horvath, D. S4MPLE-sampler for multiple protein-ligand entities: simultaneous docking of several entities. *J Chem Inf Model* **2013**, *53*, 88-102.

50. Filippakopoulos, P.; Knapp, S. The bromodomain interaction module. *FEBS Lett* **2012**, *586*, 2692-2704.

51. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. Acta *Crystallogr D Biol Crystallogr* **2010**, *66*, 133-144.

52. Kabsch, W. XDS. Acta *Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.

53. Collaborative Computational Project, N. m. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 760-763.

54. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. Acta *Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501.

55. Nicholls, R. A. Ligand fitting with CCP4. *Acta Crystallogr D Struct Biol* **2017**, *73*, 158-170.

Table 1

Compound[a]	VS Rank[b] (top %)	pIC_{50}[d]	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[c] (1.7)	6.2±0.01	1.3
3	10[c] (1.7)	6.2±0.03	1.3
4	30 (5.2)	6.0±0.02	1.1
5	40[c] (6.9)	6.3±0.02	1.4
6	40[c] (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[c] (10.2)	6.4±0.02	1.5
10	59[c] (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

[a] *Compounds discussed in the text, see **Fig. S2** for 2D structures of these compounds.*
[b] *Absolute ranking position after virtual screening and relative position in parenthesis.*
[c] *Several compounds show the same interaction energy value and are given the same rank.*
[d] *pIC_{50} 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	ΔTm (°C)[a]	K_D (μM)[b]
1	1.50±0.00	na[c]
2	3.50±0.00	0.59±0.30
3	3.00±0.00	0.56±0.13
4	3.00±0.00	0.94±0.05
5	2.88±0.25	0.32±0.09
6	2.63±0.25	0.54±0.09
7	3.13±0.25	0.36±0.09
8	2.13±0.25	0.86±0.27
9	3.00±0.00	0.31±0.05
10	3.38±0.25	0.41±0.09
13	3.00±0.00	0.77±0.12
14	3.25±0.29	0.89±0.16
17	3.88±0.25	0.19±0.07

[a] *ΔTm values were measured using a thermal shift assay in triplicates.*
[b] *K_D values were measured by ITC in triplicates.*
[c] *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 μM 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® Laboratory Workstation to be evaluated using homogeneous time-resolved fluorescence (HTRF®) 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

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

Figure 3. Experimental evaluation of compound 17. A. 2D structure of compound **17** (4-[(2-oxo-2,3,4,5-tetrahydro-1H-1-benzazepine-7-sulfonamido)methyl]-3-benzyl-7-ethyl-1H-purine-2,6(3H,7H)-dione)). **B.** Homogeneous time-resolved fluorescence (HTRF) of **17** on BRD4(BD1) after excitation at 337 nm, energy transfer at 620 nm, and fluorescence emission at 665 nm. Fluorescence data are normalized and plotted as a function of the ligand concentration. The IC_{50} value is indicated in nanomolar. **C.** Isothermal titration calorimetry (ITC) thermogram (top) and nonlinear least-squares fit model of the integrated data (bottom) at 15°C (100 µM BRD4(BD1) *vs.* 40 µM compound **17**) as a function of the molar ratio of the ligand (cell) to the protein (syringe). The K_D value is indicated in nanomolar.

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

Å (PDB code: 6FNX) and 0.95 Å (PDB code: 6FO5) resolution, respectively. Ligands are displayed in ball-and-stick representation. Key residues of BRD4(BD1) are represented as sticks and colored green for the WPF shelf (residues 81-83), orange for Y97, pink for N140, blue for the gatekeeper I146 and white for the others. Water molecules are represented by red spheres. Direct hydrogen bonds between the ligand and protein are represented by gray dashed lines. **D-E.** The network among water molecules located in the cavity is represented by red dashed lines. The indirect bonds between ligand and protein mediated by water molecules are represented by orange dashed lines. **F-G.** Van der Waals contacts between the ligand and protein are represented by light-blue dashed lines.

Figure 1

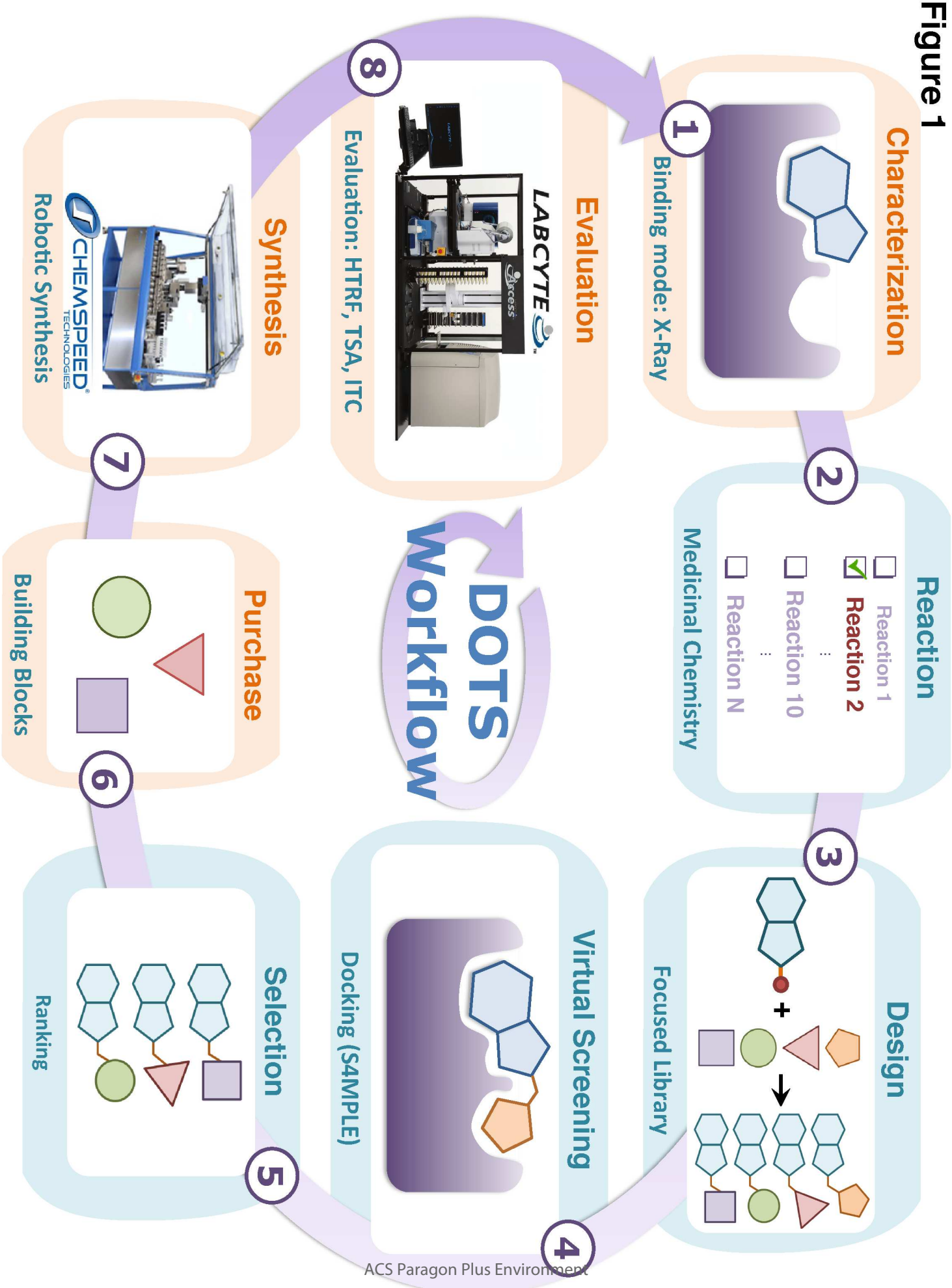

Characterization
1
Binding mode: X-Ray
2
Reaction
Reaction 1
Reaction 2
...
Reaction 10
...
Reaction N
Medicinal Chemistry
3
Design
Focused Library
4
Virtual Screening
Docking (S4MPLE)
5
Selection
Ranking
6
Purchase
Building Blocks
7
Synthesis
Robotic Synthesis
CHEMSPEED TECHNOLOGIES
8
Evaluation
Evaluation: HTRF, TSA, ITC
LABCYTE
Excess
DOTS Workflow

Figure 2

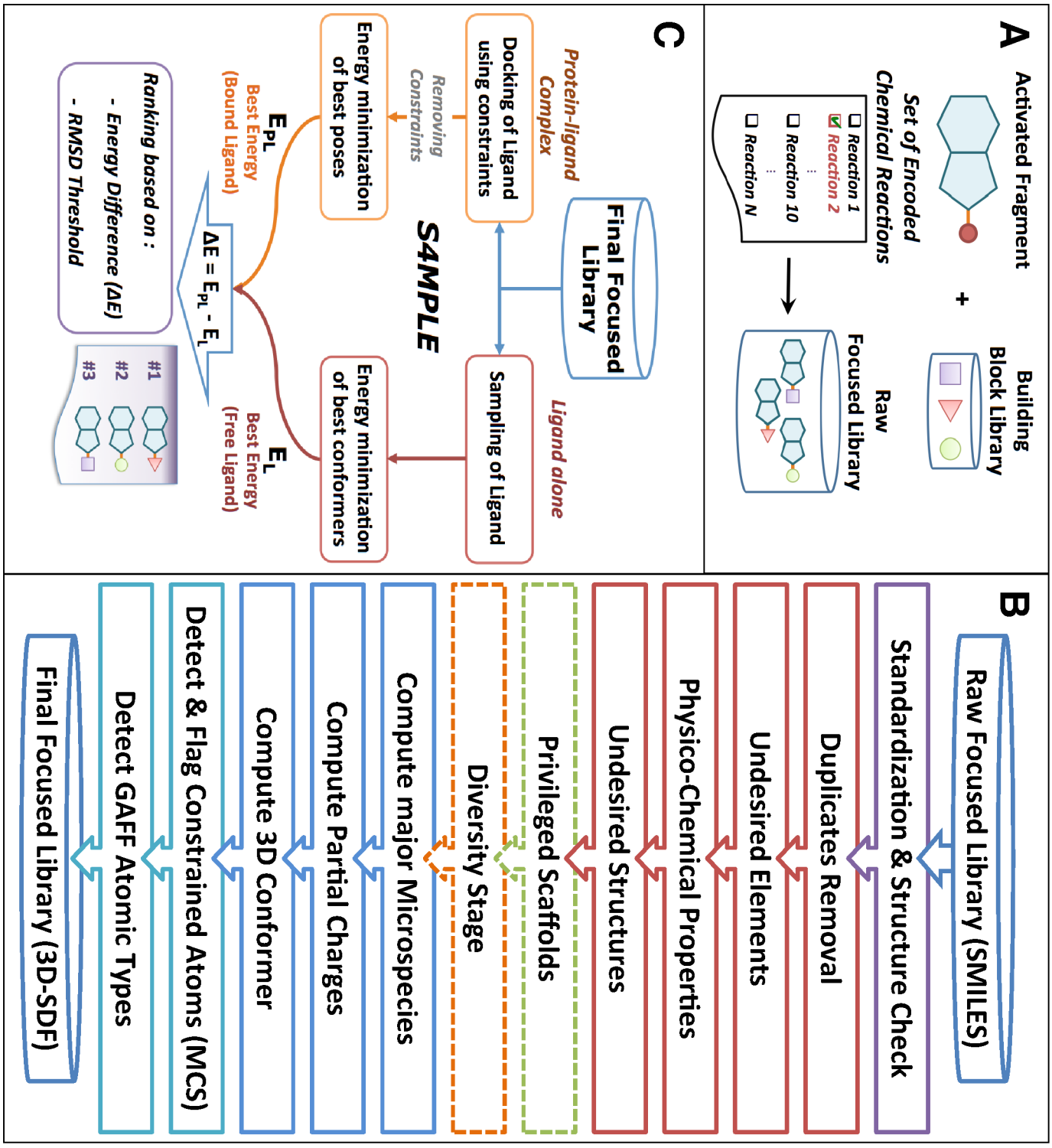

A
Activated Fragment
+
Building Block Library
Set of Encoded Chemical Reactions
Reaction 1
Reaction 2
Reaction 10
Reaction N
Raw Focused Library
B
Raw Focused Library (SMILES)
Standardization & Structure Check
Duplicates Removal
Undesired Elements
Physico-Chemical Properties
Undesired Structures
Privileged Scaffolds
Diversity Stage
Compute major Microspecies
Compute Partial Charges
Compute 3D Conformer
Detect & Flag Constrained Atoms (MCS)
Detect GAFF Atomic Types
Final Focused Library (3D-SDF)
C
Protein-ligand Complex
Docking of Ligand using constraints
Removing Constraints
Energy minimization of best poses
Final Focused Library
S4MPLE
Ligand alone
Sampling of Ligand
Energy minimization of best conformers
E_PL Best Energy (Bound Ligand)
ΔE = E_PL - E_L
E_L Best Energy (Free Ligand)
Ranking based on :
- Energy Difference (ΔE)
- RMSD Threshold
#1
#2
#3

Figure 3

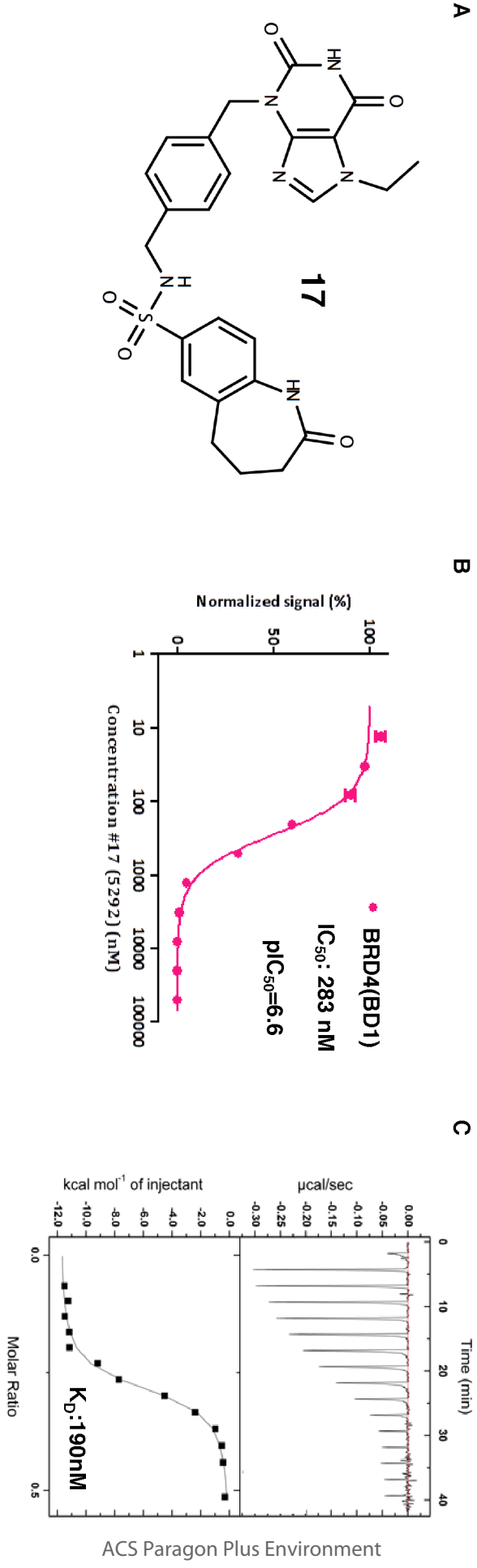
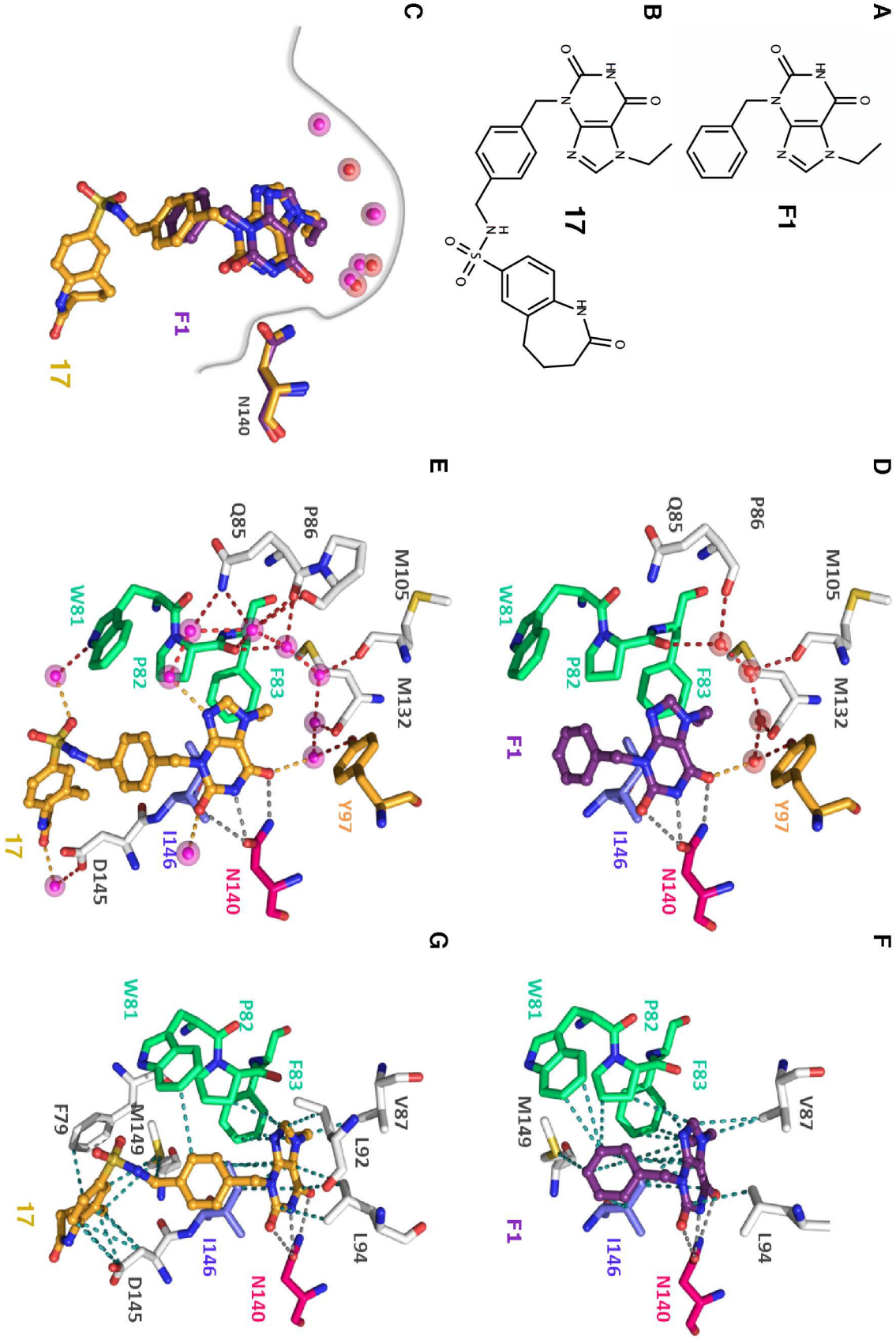

A
17
B
Normalized signal (%)
Concentration #17 (5292) (nM)
BRD4(BD1)
IC50: 283 nM
pIC50=6.6
C
kcal mol-1 of injectant
µcal/sec
Time (min)
Molar Ratio
KD:190nM

Figure 4


A
F1
B
17
C
F1
17
N140
D
M105
Q85
P86
W81
P82
F83
M132
F1
Y97
I146
N140
E
M105
Q85
P86
W81
P82
F83
M132
Y97
17
I146
D145
N140
F
W81
P82
F83
V87
M149
F1
I146
L94
N140
G
W81
P82
F83
V87
M149
F79
L92
17
I146
L94
D145
N140