JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Discovery of Inhibitors of Four Bromodomains by Fragment-Anchored Ligand Docking

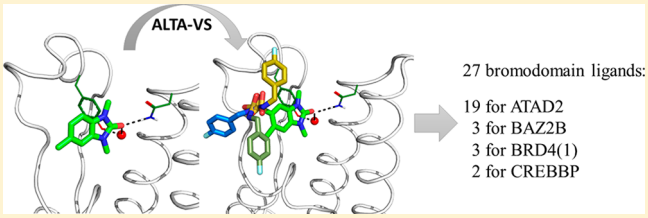
Jean-Rémy Marchand,[†] Andrea Dalle Vedove,[‡] Graziano Lolli,[‡] and Amedeo Caflisch*,[†]

[†]Department of Biochemistry, University of Zürich, CH-8057, Zürich, Switzerland

[‡]Centre for Integrative Biology, University of Trento, I-38123, Povo, Italy

Supporting Information

ABSTRACT: The high-throughput docking protocol called ALTA-VS (anchor-based library tailoring approach for virtual screening) was developed in 2005 for the efficient in silico screening of large libraries of compounds by preselection of only those molecules that have optimal fragments (anchors) for the protein target. Here we present an updated version of ALTA-VS with a broader range of potential applications. The evaluation of binding energy makes use of a classical force field with implicit solvent in the continuum dielectric approximation. In about 2 days per protein target on a 96-core compute cluster (equipped with Xeon E3-1280 quad core processors at 2.5 GHz), the screening of a library of nearly 77 000 diverse molecules with the updated ALTA-VS protocol has resulted in the identification of 19, 3, 3, and 2 μM inhibitors of the human bromodomains ATAD2, BAZ2B, BRD4(1), and CREBBP, respectively. The success ratio (i.e., number of actives in a competition binding assay in vitro divided by the number of compounds tested) ranges from 8% to 13% in dose−response measurements. The poses predicted by fragment-based docking for the three ligands of the BAZ2B bromodomain were confirmed by protein X-ray crystallography.



INTRODUCTION

In vitro fragment-based drug design is an attractive strategy to cover the chemical space of binders at a reduced cost and a more efficient approach than brute force high-throughput screens.[1] On the computational side, high-throughput docking is a valuable asset when it comes to finding small molecule binders of proteins.[2,3] It can enrich screening libraries usually with 1% to 10% hit rates.[4,5] Recently, evidence has accumulated on the success of fragment−hit identification with force field-based approaches which make use of explicit[6] or implicit solvent treatment.[5,7−9]

In 2005, an efficient computational approach that combines the advantages of high-throughput docking with those of fragment-based hit identification was introduced.[10,11] The computational protocol was later called anchor-based library tailoring approach virtual screening (ALTA-VS, Figure 1).[12] ALTA-VS is a four-step protocol: (1) decomposition of the chemical library into its essential rigid fragments; (2) docking of the fragments and evaluation of binding energy (with generalized Born approximation of electrostatic solvation effects); (3) flexible docking of the parent molecules that contain the top ranking fragments which are used as noncovalent binding anchors during docking (three anchor fragments for each parent molecule in the original version of ALTA-VS); and (4) energy minimization with final evaluation of binding energy including desolvation effects in the continuum dielectric approximation (finite-difference Poisson equation). An essential element of the ALTA-VS approach is the much higher efficiency of the fragment-anchored docking of a set of 10^3 to 10^4 parent molecules than the docking of a multimillion compound library and, importantly, the high accuracy of docked poses of fragments.[5,7−9] The successful identification of hit compounds by the ALTA-VS approach has been reported for several protein targets.[10−14]

Here we report on an updated version of the ALTA-VS method, which has three improvements with respect to the original protocol.[10,11] First, a single anchor fragment is required for flexible ligand docking instead of three fragments as in the original version of ALTA-VS. Second, it employs an energy function without any fitting parameters. These two improvements result in a broader range of applicability, in particular to protein targets with a small binding site and/or for which inhibitors have not been disclosed. Third, we make use of a transferable force field which treats in a consistent way the parameters of proteins (CHARMM36)[15] and organic compounds (CGenFF).[16,17]

Bromodomains are left-handed four-helix bundles of about 110 residues.[18] The 61 human bromodomains, found in 46 proteins, bind acetylated peptides and in particular acetylated histone tails.[19] The bromodomain binding pocket has been heavily investigated, both in complex with physiological ligands or with synthetic small molecules.[20] The binding pocket of the natural ligand (acetylated lysine in histone tails and other proteins) is formed on one of the two ends of the four-helix bromodomain fold by the loop connecting the helix αZ and αA,

Received: June 6, 2017
Published: September 1, 2017

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

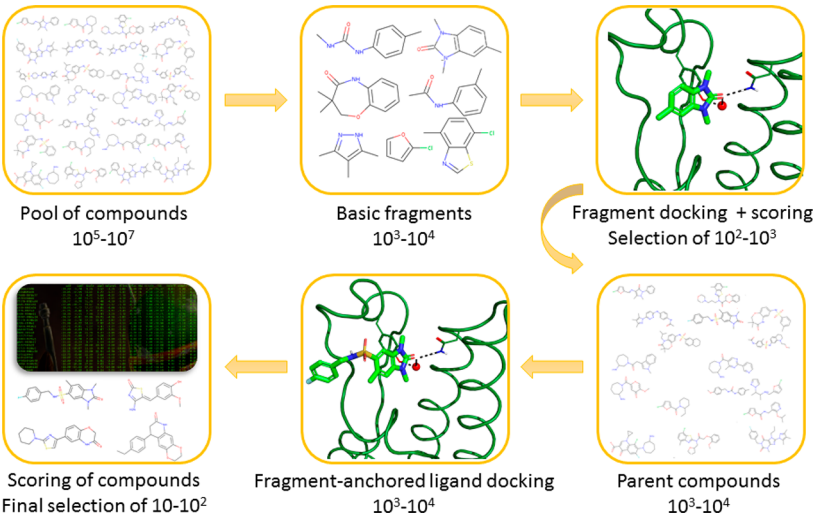


Figure 1. Anchor-based library tailoring approach for virtual screening (ALTA-VS). A chemical library, with up to tens of millions of compounds, is decomposed in nonrotatable fragments, which are docked and scored. Parent compounds containing the top ranking fragments are retrieved and docked with tethering of the fragment headgroup. Those docked molecules are then further energy minimized with a force field and evaluation of electrostatic desolvation effects by the finite-difference Poisson approach. Thus, the ALTA-VS protocol selects $10-10^2$ compounds for in vitro validation (bottom, left panel) from libraries of 10^5-10^7 molecules (top, left) by docking only 10^3-10^4 fragments (top, middle) and 10^3-10^4 compounds (bottom, middle).

called ZA loop, and the loop linking the αB and αC helices, named the BC loop. The bromodomain family has high therapeutic interest, with 14 inhibitors currently in clinical trials.[21,22] Their crucial implication in epigenetics affects gene expression with potential involvement in cancer, inflammation, and neurological diseases.[23]

The new version of ALTA-VS is applied here to screen a library of nearly 77 000 compounds for the bromodomains of ATAD2, BAZ2A, BAZ2B, BRD4(1), and CREBBP. In contrast to previous in silico screening campaigns which focused each on a single protein target and made use of different libraries,[5,24,25] the same protocol (updated ALTA-VS) and library were employed in the present study for five related proteins. The updated ALTA-VS protocol has led to the discovery of 27 small-molecule inhibitors with micromolar potency for the target bromodomain and, in particular, 15 compounds with an affinity below 100 μM. Furthermore, the predicted binding mode of the three inhibitors of the BAZ2B bromodomain was confirmed by protein X-ray crystallography.

■ THEORY

The ALTA-VS strategy consists of four steps here illustrated in detail. The first step is the decomposition of each molecule of the library into rigid fragments, which are defined such that all rotatable bonds are cut. The original library is usually a collection of commercially available compounds. In contrast, the generated fragment library includes molecules that may not be synthesizable, which is not a limitation as the final ranking and selection for experimental validation are done for the original compounds (Figure 1).

In the second step, the fragment library is docked with the program SEED[26,27] which requires about one second per fragment.[5] SEED docks the rigid fragments in the target binding pocket by an exhaustive search with intermolecular hydrogen bond distances as restraints. For each fragment position and orientation the binding energy is evaluated with a force field and the generalized Born implicit solvation model is used to approximate the electrostatic contribution to the binding free energy.[28]

For the third step, the fragments with the most favorable calculated binding energy are selected. Their parent compounds are retrieved and docked with a flexible docking tool. Here we used RDock,[29] which is adequate for the ALTA-VS approach as it can dock compounds with a tethered substructure. For each parent compound, the substructure corresponding to the docked fragment is tethered in the binding pocket and the orientation of the rest of the molecule is optimized by a search in the space of rotatable bonds. Flexible docking software usually employ a very crude approximation of the binding free energy, especially for the solvation and torsional energies. Those approximations can lead to implausible binding poses, which result in a large amount of false positives in the top ranked compounds. Moreover, previous reports have emphasized the difficulties in docking highly flexible compounds.[30] In the ALTA-VS approach the size of the conformational space of flexible ligands is reduced substantially by fragment anchoring. The original implementation of the ALTA-VS procedure[12] made use of the SEED docking of fragments in three different subpockets and subsequent linkage of the fragments with a flexible docking algorithm developed in-house called FFLD.[31] This approach required the empirical definition of three subpockets, which might not be possible for targets with small and/or shallow binding site. For instance, one can hardly define three subpockets in the binding site of bromodomains.[32] The availability of open source software for flexible docking with possibilities of substructure tethering such as RDock gave us the opportunity to sample docking positions of putative ligands with a single anchored headgroup.

Once the molecules are docked, the fourth and last step consists in minimizing the energy of the poses and rescoring them, i.e., estimating their binding free energy. Here, we employ the Poisson equation solver[33] of the CHARMM package[34] to compute the electrostatic component of the binding process, including solvation energy, and add the

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

electrostatic term to the ligand/bromodomain van der Waals energy. Importantly, a force field-based energy function was used here without any fitting parameters, while the original ALTA-VS protocol used a scoring function based on a linear interaction energy model with two multiplicative parameters for van der Waals and electrostatics contributions, respectively.[12]

■ EXPERIMENTAL SECTION

Assembly of the Screening Library. The chemical library[35] of the Lausanne Bioscreening Facility was used here because it is a large collection of diverse compounds which are commercially available. The library we assembled from their service contained 76 731 chemically diverse molecules from seven vendors (see Supporting Information, section S1).

We used the ChemAxon software suite to prepare the library, with a special focus on protonation and tautomeric states. To the best of our knowledge and at the time of the study, the ChemAxon software suite relied on the largest training set for organic molecule pK_A calculations.[36] The electronic library was processed and cleaned with the Calculator Plugins of Marvin 15.1.5.0, 2015, ChemAxon (www.chemaxon.com) with the following steps:

- ChemAxon structurechecker module to check for aromaticity and valence issues.
- ChemAxon stereoisomers module to generate a maximum of four stereoisomers per molecule.
- ChemAxon dominanttautomerdistribution module to generate tautomers (protomers) at pH 7.2, filtering out structures with a predicted occupancy below 25%.
- ChemAxon leconformer module to generate one clean conformer per molecule (optimization level 3, MMFF94, in vacuo).

This preparation protocol generated 135 840 structures, corresponding to 75 830 unique molecules. Thus, the loss of molecules was about 1%. Chemical library properties, e.g., molecular weight, number of hydrogen bond donors, heavy atom count, are available in the Supporting Information, section S1.

Fragmentation of the Library. The molecules were fragmented into their essential rigid fragments by cutting all rotatable bonds with a script relying on the RDKit.[37] Rotatable bonds were defined according to the following SMARTS pattern: [!$([NH]!@C(=O))&!D1&$(*#*)]&!@[$([NH]!@C(=O))&!D1&!$(*#*)]. Missing valences were filled with a hydrogen atom, which only for a small number of fragments led to spurious hydrogen-bond donors. The generated virtual fragments were subsequently uniquified with Open Babel,[38] which led to 6436 rigid fragments. In a next step, we prepared the fragments for docking with SEED. The software CGenFF 1.0 was used to assign CGenFF 3.0.1 parameters to the fragments.[16,17] After parametrization and preparation, 6406 fragments were left for docking, which corresponds to a loss of only 0.5% of the fragments during the parametrization.

Preparation of Protein Targets. Crystal structures of proteins were selected with a preference for those solved in the laboratory of the authors and according to the following rationale: (1a) does an in-house holo crystal structure exist? Yes for BAZ2B (PDB code 5e73), BRD4(1) (4pci), and CREBBP (4tqn). (1b) If not, is there a holo crystal structure available in the literature? Yes for ATAD2 (5a5r) and BAZ2A (4bqm). (2) We chose the structure with the most potent small organic binder at the time of the study (in-house for 1a and in the PDB for 1b), with the exception of BAZ2A for which a single holo structure was available bound to an acetylated histone tail peptide. Hydrogen atoms of the protein and the six structurally conserved water molecules[19] were added and minimized with CHARMM.[34] The minimization took place in the presence of the ligand of the crystal structure. Heavy atoms were fixed, while added atoms were free to move during a two-stage minimization in vacuo which consisted of 5000 steps of steepest descent and 100 000 steps of conjugate gradient with convergence threshold based on the energy gradient of 0.01 kcal/(mol Å). The convergence threshold was reached in all case.

Docking and Selection of Anchor Fragments. Docking of fragments was carried with our in-house docking software SEED[26,27] with the same protocol for the five bromodomains. The six conserved water molecules were considered part of the receptor. The binding site was defined by two residues and two structurally conserved water molecules. The two residues are the evolutionary conserved asparagine whose side chain is involved in a hydrogen bond with the acetyl group of the natural ligand (Asn1064 for ATAD2, Asn1873 for BAZ2A, Asn1944 for BAZ2B, Asn140 for BRD4(1), Asn1168 for CREBBP), and the central residue of the so-called WPF shelf (Val1008 for ATAD2 which has an RVF shelf, Pro1817 for BAZ2A, Pro1888 for BAZ2B, Pro82 for BRD4(1), Pro1110 for CREBBP). The two water molecules are the one that bridges the acetyl group to the conserved tyrosine (Tyr1021 for ATAD2, Tyr1830 for BAZ2A, Tyr1901 for BAZ2B, Tyr97 for BRD4(1), Tyr1125 for CREBBP) and the conserved water molecule the farthest from it (HOH3081 for ATAD2, HOH2214 for BAZ2A, HOH2138 for BAZ2B, HOH381 for BRD4(1), HOH1339 for CREBBP). The interior dielectric constant of the protein was set to 2.0, and the solvent dielectric constant, to 78.5.

Following previous evidence of the importance of key hydrogen bonds as a filter,[13] we discarded the fragments that did not have any hydrogen bond-acceptor atom within a distance of 4 Å from the side chain nitrogen of the conserved asparagine. Three energy-based functions were used to filter compounds from the different energy components calculated by SEED: Delta electrostatics, total energy efficiency, and electrostatic efficiency. The Delta electrostatics is the difference between the fragment/receptor electrostatic interaction and the electrostatic free energy of hydration of the fragment (both calculated with the generalized Born approach). It measures the intermolecular electrostatic interaction relative to the free energy of solvation of the fragment assuming a rigid conformation of both protein and fragment. The total energy efficiency is the total binding free energy divided by the number of non-hydrogen atoms (HAC = heavy atom count), which tend to penalize large ligands. The electrostatic efficiency is the fragment/receptor electrostatic interaction divided by HAC. This term tends to favor compounds with favorable electrostatic interaction with the protein, normalized by the number of heavy atoms to avoid a systematic bias toward large polar compounds.

For each of the five bromodomains, the top 150 fragments were selected individually based on Delta electrostatics, total energy efficiency, and electrostatic efficiency. Then, the fragment selections between targets were compared to be mutually exclusive for each target. In other words, only fragments selected for a single bromodomain were retained, i.e.,

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Statistics of the ALTA-VS Protocol in the Five Bromodomains

	ATAD2	BAZ2A	BAZ2B	BRD4(1)	CREBBP
number of fragments (out of 6406)	142	137	184	210	192
unique parent compounds	2954	894	2791	2520	2476
conformers generated by the RDKit	21831	9095	22408	20456	20521
conformers tethered by RDock	19861	5122	15255	10269	10209
poses obtained by docking	393620	102440	305100	205380	204180
poses minimized and scored after clustering (of X unique compounds)	202541 (2819)	85889 (636)	168680 (1997)	114293 (1290)	105264 (1254)

promiscuous fragments were discarded. This resulted in a number of fragment ranging from 137 (for BAZ2A) to 210 (for BRD4(1), Table 1). Interestingly, the numbers of fragments passing the filters reflect the druggability of the related bromodomains: BRD4(1) is considered the most promiscuous of the 61 human bromodomains, followed by CREBBP.[39] BAZ2B is usually considered a difficult target,[40] but the fast growing number of binders[7,41] would tend to rank it as a midrange druggability target. The bromodomains of BAZ2A and ATAD2 have been reported as very difficult targets.[42] They were also the two bromodomains with the smallest number of fragments identified by ALTA-VS.

Docking of Parent Compounds. The number of parent compounds of the top ranking fragments ranged from 894 (for BAZ2A) to 2954 (for ATAD2) (Table 1). Docking was then carried out with RDock, a flexible docking software by tethering the fragment in the pose it had after SEED fragment docking. Diverse conformers of the parent compounds were generated with RDKit, using the ETKDG algorithm.[43] Fifty conformer generation runs were used per molecule and a threshold of diversity of 2 Å between conformers. Tethering was performed with the helper script of RDock, which relies partly on Open Babel, leading to some loss during the automatic detection of substructures in molecules. For flexible ligand docking, the binding site definition made use of a radius of 10 Å around each atom of the crystal ligand for all bromodomains except BAZ2A. Since BAZ2A had an acetylated peptide as ligand, we used the ligand of the closest analog, BAZ2B, after structural superposition. During the docking runs, the conserved water molecules were kept fixed and the compounds' head groups tethered as explained earlier. Other parameters were set as default. Twenty docking runs were carried out for each bromodomain/small molecule pair.

Rescoring and Selection of Compounds. Poses were clustered to avoid redundant binding modes with the RDKit (threshold of 1 Å) and converted to CHARMM file formats. Statistics on number of poses and compounds are in Table 1. A total of 676 667 poses were energy minimized by CHARMM which was used also to evaluate the electrostatic contribution to the binding free energy by finite-difference Poisson calculations. During minimization the bromodomain atoms and six conserved water molecules were kept rigid. The minimization protocol consisted of 500 steps of steepest descent followed by 10 000 steps of conjugate gradient, with a tolerance of the energy gradient of 0.01 kcal/(mol Å). A distance-dependent dielectric constant of $4r$ (where r is the distance between atomic nuclei, i.e., positions of partial charges) was used during the minimization to avoid in vacuo minimization artifacts.[44,45] The CHARMM pbeq module was used for the finite-difference Poisson calculations on the minimized structures.[33] The dielectric constant was set to 4.0 for the solute (bromodomain, six structural waters and ligand) and to 78.5 for the solvent. The dielectric constant of 4.0 for the solute is a factor of 2.0 larger than the one used for docking the fragments (in the generalized Born calculations in SEED). It was chosen to partially account for the flexibility of the ZA loop, which is more relevant for the docking of the parent compounds than the head-groups. The grid for the Poisson equation calculation was centered on the center of mass of the protein, with a nonfocused grid spacing of 1.0 Å and a focused grid spacing of 0.3 Å. The number of grid points was automatically calculated in each dimension as $(40 + (X, Y, Z)_{max} - (X, Y, Z)_{min})/1.0$ for the nonfocused grid and $(10 + (X, Y, Z)_{max} - (X, Y, Z)_{min})/0.3$ for the focused grid. The 40 and 33 (= 10/0.3) additional grid points in each dimension for the unfocused and focused grid, respectively, are required to extend the boundary by 20 and 5 Å. The ionic strength was set to zero as it has little influence on the single point calculations of the binding free energy of protein−ligand complexes typical in pharmaceutical applications.[46−49] In the absence of an ionic atmosphere the Poisson−Boltzmann equation reduces to the Poisson equation. Similarly, no term was included for the nonpolar contribution to the free energy of solvation.[50−54]

After energy minimization, some docked poses could move out of the asparagine subpocket because of unrealistically strained conformations or strongly unfavorable contacts. A filter was applied again to keep only compounds that interact with the conserved asparagine (distance <4.0 Å to the side chain nitrogen atom). Upon filtering, the top 100 compounds according to Delta electrostatics, and top 100 according to total energy efficiency were selected for each bromodomain. Some of these compounds were discarded as they contained chemotypes of known bromodomain inhibitors (e.g., dimethylisoxazole and acetylbenzene derivatives), or were commercially unavailable (in January 2016). Finally, 142 molecules (out of 186) were selected for testing for ATAD2, 30 (of 65) for BAZ2A, 25 (of 67) for BAZ2B, 38 (of 152) for BRD4(1), and 25 (of 163) for CREBBP (Table 2). Only a fraction of compounds were selected for the bromodomains of BRD4(1) and CREBBP as many of the top ranking compounds showed some similarity with known inhibitors of these two highly targeted bromodomains. Furthermore, for four of the five bromodomains several compounds were discarded because of close similarity among each other (e.g., pairs of molecules differing by a single halogen atom) while for ATAD2 almost all top ranking compounds were purchased because of the low druggability of this target and to generate some initial SAR. All the scoring and experimental data of the purchased compounds are available in the Supporting Information, section S5, including smiles chemical format depiction.

Validation Assays. Competition binding assays were performed to verify which compounds were true binders of their respective targets. Molecules were first tested in a single dose experiments at high concentration depending on the

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Summary of Experimental Results Per Target[a]

	ATAD2	BAZ2A	BAZ2B	BRD4(1)	CREBBP	total
N tested	142	30	25	38	25	260
active in single dose	44	3	10	9	2	68
IC_{50}/K_D < 400 μM	19	0	2	3	2	26
IC_{50}/K_D < 100 μM	10	0	2	2	1	15
X-ray structures	N/A	1	2	N/A	N/A	3

[a]Actives in single dose were defined by using a threshold of less than 65% inhibition of the competitor ligand. Values of IC_{50} were measured for ATAD2, BRD4(1), and CREBBP by AlphaScreen, while K_D values were measured for BAZ2A and BAZ2B by BROMOScan. X-ray crystallography was carried out only for the BAZ2A and BAZ2B bromodomains.

solubility (between 50 μM and 500 μM). In all cases, the DMSO concentration was 0.1%, except for ATAD2 for which it was 1%. ATAD2, BRD4(1), and CREBBP experiments were carried out by AlphaScreen assays[55] at Reaction Biology Corp. The AlphaScreen assay makes use of a donor bead (on the competitor molecule) that can transfer singlet oxygen to an acceptor bead (on the bromodomain target) when the two beads are in close proximity (<200 nm). The acceptor bead then emits a luminescent signal. When a compound binds the target, the donor/acceptor complex is disrupted, leading to a loss of singlet oxygen transfer and loss of the signal. The donor was histone H4 peptide (1−21) K5/8/12/16Ac-Biotin for ATAD2, CREBBP, and BRD4(1). The BAZ2A and BAZ2B assays were performed at DiscoverX with the BROMOscan profiling service.[56] The BROMOscan assay is a competition binding assay based on DNA-tagged bromodomain and quantitative PCR. The ligand used for the assay is proprietary and undisclosed.

The 68 compounds exhibiting a remnant binding of the competitor below 65% with respect to DMSO solution (Table 2) were considered for dose−response assays, i.e., determination of IC_{50} (ATAD2, BRD4(1), and CREBBP at Reaction Biology Corp.) and K_D (BAZ2A and BAZ2B at DiscoveRX) values. Out of these 68 molecules, the number was reduced to 39 mostly because of compound solubility, chemical novelty, and cost management. The binding affinity were investigated with the same assays as described before, with a curve fitting of 10-point dose responses for the AlphaScreen assay (ATAD2, BRD4(1), and CREBBP) and a curve fitting of 12-point dose responses in duplicates for the BROMOscan assay (BAZ2A, BAZ2B) (dose−response curves in the Supporting Information, section S3). Positives controls were as follows: JQ1[57] for ATAD2 (IC_{50} = 64 μM) and BRD4(1) (IC_{50} = 0.02 μM), SGC-CBP30[58] for CREBBP (IC_{50} = 0.08 to 0.15 μM), and undisclosed for BAZ2A and BAZ2B.

It is important to provide evidence of specific binding for the active compounds. The 26 compounds that showed activity in dose−response assays were negative in tests for known pan-assay interference (PAINS) substructures at the FAF-Drugs4 Web server.[59] Furthermore, the 26 actives resulted negative in a test for known aggregation substructures or properties at the Aggregator Advisor Web server.[60]

X-ray Crystallography. BAZ2A and BAZ2B bromodomains were produced and crystallized as described previously.[61] Briefly, proteins were purified by IMAC, followed by buffer exchange, tag removal by TEV protease, a second IMAC and a size-exclusion chromatography. Complexes with the compounds of interest were obtained by cocrystallization for BAZ2A and by soaking for BAZ2B. Compounds were dissolved in the crystallization solution devoid of DMSO and MPD, which bind to the binding pocket of bromodomains.[62] Diffraction data were collected at the Elettra Synchrotron Light Source (Trieste, Italy), XRD1 beamline (PDB codes 5OR8 and 5OR9) and at the Swiss Light Source, Paul Scherrer Institute (Villigen, Switzerland), beamline PXI (PDB code 5ORB). Data were processed with XDS[63] and Aimless;[64] high resolution cutoff was selected according to Karplus and Diederichs.[65] BAZ2B crystals soaked with compound 30 were strongly anisotropic: ellipsoidal diffraction data were treated making use of the STARANISO server (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). Structures were solved by molecular replacement with Phaser[66] using PDB 4IR5 as search model for BAZ2B and PDB 5MGJ for BAZ2A. Initial models were refined alternating cycles of automatic refinement with Phenix[67] and manual model building with COOT.[68]

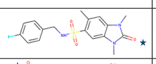
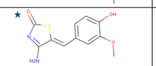
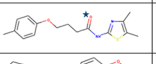
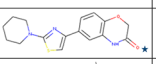
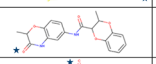
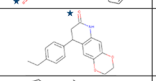
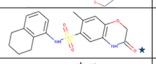
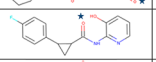
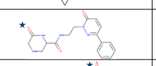
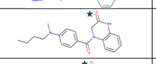
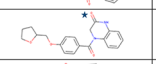
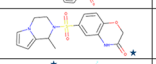
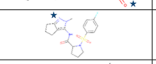
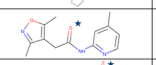
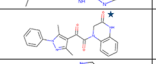
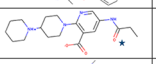
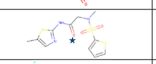
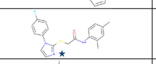
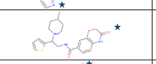
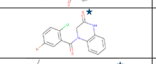
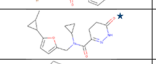
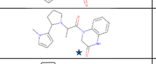
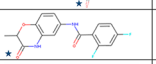
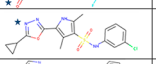
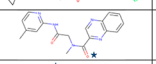
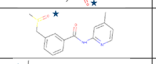
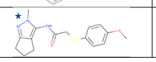
■ RESULTS AND DISCUSSION

ALTA-VS Identifies Binders for Four of the Five Bromodomains. The binding free energy of nearly 700 000 poses was evaluated upon minimization and finite-difference Poisson calculation of the electrostatic contribution to desolvation (see the Experimental Section). The top 25−38 compounds were selected for 4 of the 5 bromodomains, while 142 compounds were retained for ATAD2, which is a very difficult target.[69] Evaluation of the 260 compounds was performed by competition-binding assays (AlphaScreen at Reaction Biology Corp. and BROMOscan at DiscoverX). A total of 68 compounds (26%) showed less than 65% binding of the competitor molecule in single dose experiments at maximum dilution concentration (between 50 and 500 μM for most compounds). Thirty-nine of the 68 single-dose actives were evaluated in dose−response assays. The remaining 29 molecules were not considered further because of solubility and/or novelty considerations. Of the 39 single-dose actives that underwent dose−response measurements, 26 compounds showed an affinity below 400 μM, and 15 compounds below 100 μM (Table 2). Compound **30** showed significant inhibition of the BA2B bromodomain at a single dose of 50 μM but did not show activity in the dose−response assay. It is a competitive binder as its crystal structure in complex with BAZ2B confirmed that it binds in the pocket of the natural ligand (see below). The definition of an affinity threshold is not straightforward. It has been reported by several groups that the K_D values measured by the BROMOscan assay are up to an order of magnitude more favorable than the IC_{50} values obtained by AlphaScreen.[5,70−72] We decided to use the same threshold for both assays because most of the compounds (33 of 39) were evaluated by the Alphascreen assay. If one excludes BAZ2A (see below), the success ratio (number of actives/number of tested molecules) ranged from 8% to 13% (19 binders of 142 compounds tested for ATAD2) with the 400-μM threshold and from 4% to 8% (2 binders of 25 compounds tested for BAZ2B) with the100-μM threshold (Tables 2 and 3).

In addition to the 26 binders identified in dose−response measurements, we investigated the binding of compounds **1**, **13**, **30**, **31**, and **187** (Supporting Information section S4) by means of X-ray crystallography. These five compounds originate from the docking into BAZ2B and were tested by

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 3. In Vitro Results for the 26 Compounds That Showed Activity in a Dose−Response Assay and Compound 30[b]

Cpd	Target	Structure	% remnant inhibition of competitor at [μM]	IC_{50} μM	Calculated Delta electr. kcal/mol	Calculated total energy efficiency kcal/(mol HAC)	Calculated total energy kcal/mol
1	BAZ2B		1% 90 μM	6[a]	2.7	-1.1	-27.5
2	BRD4(1)		38% 200 μM	22	0.9	-1.3	-21.4
3	ATAD2		10% 350 μM	23	-2.0	-0.9	-19.5
4	ATAD2		38% 50 μM	34	-1.6	-0.8	-17.8
5	ATAD2		40% 504 μM	34	-1.3	-0.9	-22.2
6	ATAD2		48% 504 μM	35	-1.4	-1.2	-28.6
7	BRD4(1)		23% 201 μM	37	0.4	-1.1	-27.8
8	ATAD2		4% 503 μM	44	-1.7	-1.3	-26.2
9	CREBBP		0% 503 μM	55	1.4	-1.1	-28.1
10	ATAD2		48% 318 μM	67	-0.9	-0.9	-22.0
11	ATAD2		48% 200 μM	70	-0.3	-1.0	-25.8
12	ATAD2		12% 479 μM	78	-2.4	-0.8	-19.5
13	BAZ2B		18% 160 μM	88[a]	2.6	-1.2	-32.0
14	ATAD2		6% 503 μM	100	-0.3	-1.0	-17.2
15	ATAD2		29% 200 μM	100	0.4	-1.1	-31.6
16	ATAD2		37% 254 μM	151	0.9	-1.1	-28.3
17	ATAD2		42% 200 μM	151	0.4	-1.1	-22.0
18	BRD4(1)		40% 201 μM	167	-0.7	-1.6	-39.1
19	ATAD2		37% 399 μM	169	11.2	-1.2	-32.2
20	ATAD2		45% 504 μM	234	-1.1	-1.3	-26.3
21	ATAD2		47% 350 μM	248	-1.7	-0.8	-18.8
22	ATAD2		45% 350 μM	258	-0.3	-1.1	-27.7
23	ATAD2		43% 505 μM	288	0.0	-0.9	-20.1
24	CREBBP		31% 588 μM	336	-0.3	-0.9	-23.2
25	ATAD2		49% 504 μM	372	-0.5	-1.1	-27.0
26	ATAD2		49% 504 μM	396	2.3	-1.0	-20.2
30	BAZ2B		41% 50 μM	> 51	-0.3	-1.6	-35.7

[a]Affinities measured for the BAZ2B bromodomain are K_D values, while for the other bromodomains, they are IC_{50} values measured by AlphaScreen. [b]The single-dose value is the percentage of remaining binding of the competitor molecule with respect to DMSO solution; thus, smaller single-dose values reflect stronger binding of the tested molecule. The compounds are ordered according to the affinity. The star indicates the hydrogen bond acceptor that is predicted to be involved in the hydrogen bonds with the side chain of the conserved asparagine and the water molecule bridging to the conserved tyrosine. X-ray structures were solved for the complexes of compound **1** and

Table 3. continued

BAZ2A (PDB code 5OR8), compound **13** and BAZ2B (5OR9), and compound **30** and BAZB (5ORB).

crystallography in both BAZ2B and BAZ2A. These two bromodomains share a very similar binding site, the only difference being the gatekeeper residue (valine in BAZ2A, isoleucine in BAZ2B).[73] Crystal structures of three of the five molecules were determined: **1** in BAZ2A (2.4 Å resolution, PDB code 5OR8), **13** and **30** in BAZ2B (2.0 Å resolution PDB code 5OR9, and 2.1 Å resolution PDB code 5ORB, respectively). The binding of **1** to BAZ2A is essentially identical to the docking pose in BAZ2B (Figure 2, panel A). The 1,3-dimethyl benzimidazolone headgroup is tilted as compared to the docking pose, probably due to the difference in the gatekeeper residue, as described in previous studies.[61,74] The crystal pose of the 1-methyl cyclopentapyrazole headgroup of compound **13** displays similar contacts with the conserved asparagine pocket as the docking pose but is less buried (Figure 2, panel B). The rest of the molecule is differently placed in the crystal pose compared to the docking pose. This can be ascribed to one or both of the following causes. First, the electron density indicates that a single enantiomer is present, or at least largely predominant, as already observed with other compounds,[7,41] while the other enantiomer was used for docking. Second, crystal contacts may influence the binding mode of the tail group. The fluorophenyl group of compound **13** is in van der Waals contact (3.5, 4.2, and 4.4 Å) with a symmetric protein chain (Figure 2, panel C). Compound **30**, which contains the same 1-methyl cyclopentapyrazole headgroup as compound **13**, has a binding mode essentially identical in the crystal structure as in the docked pose. The main difference between docking and crystal poses lays in the position of the methoxybenzene tail, which is again in contact with a symmetric unit in the crystal pose (4.4, 5.5, and 5.5 Å contacts, Figure 2), at least in its most abundantly populated orientation. Compounds **1** and **13** were confirmed binders in dose−response measurements, unlike **30**, which presented interesting single-dose activity but was not confirmed in dose−response measurement. The starting dilution for the dose−response experiment of compound **30** was 50 μM, which is much smaller than the millimolar concentration used for soaking into BAZ2B. Furthermore, electron density is clearly defined for compounds **1** and **13**, while it strongly degrades in the region of the methoxybenzene tail for compound **30**, indicating high mobility and/or multiple conformations for this part of the molecule. Overall, the docking poses of the three compounds are similar to the crystal poses: the headgroup is correctly predicted and the main source of variability comes from the tails which are partially solvent exposed and seem not to interact specifically with the protein.

It is important to note that in vitro experiments were performed on only 0.03% (for BAZ2B and CREBBP) to 0.2% (for ATAD2) of the initial library (the nearly 77 000 compounds of the Lausanne Bioscreening Facility). In four out of the five targets, micromolar binders were identified. For the BAZ2A bromodomain, the three compounds that showed activity in the single-dose measurement were not confirmed by the dose−response assay. Out of any consideration of druggability for this target, we discovered a posteriori a nonoptimal preparation of the BAZ2A bromodomain structure. The crucial water molecule that acts as hydrogen-bond bridge

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

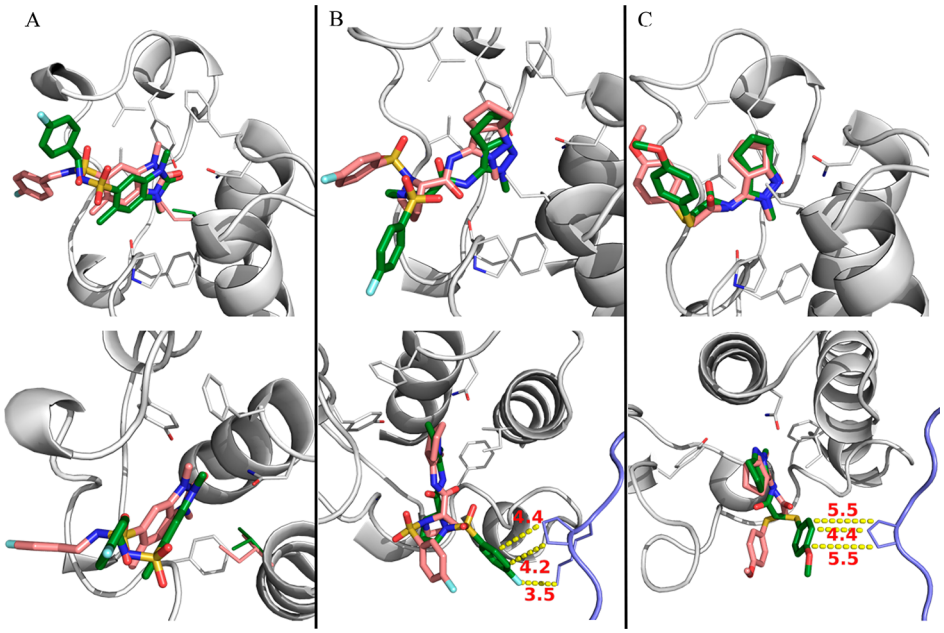


Figure 2. Comparison of binding modes from docking into BAZ2B (carbon atoms of ligands in salmon) and crystallography (carbon atoms of ligands in green). Two orientations are shown for each complex (top and bottom panels, respectively). van der Waals contacts (yellow dashed lines) with adjacent protein chains (blue) are labeled with the distance in angstroms (red). (A) Binding mode of compound **1**. The crystal pose has been solved in complex with BAZ2A (PDB code 5OR8), while the docking pose originates from BAZ2B. The major difference in the binding pocket of BAZ2A versus BAZ2B is the gatekeeper residue: valine (green) in BAZ2A and isoleucine (salmon) in BAZ2B. The difference in bulkiness between these residues explains the tilt between head-groups observed in the bottom panel. (B) Binding mode of compound **13** in BAZ2B. A different stereoisomer has been docked than the one crystallized in BAZ2B (PDB code 5OR9), which is also involved in crystal contacts (bottom part). The headgroup of the docking pose is slightly less buried than in the crystal structure. (C) Binding mode of compound **30** in BAZ2B. The crystal (PDB code 5ORB) and docking binding modes are essentially identical, with the main difference laying in the methoxybenzene tail, which contacts an adjacent protein chain of a crystal symmetric unit and is affected by a significant anisotropy (bottom part). All images were rendered using PyMOL, version 1.8.4, Schrödinger LLC.

to the side chain of the conserved tyrosine (Tyr1830) was protonated such that it could not allow for the water-mediated hydrogen bond (Figure 3). The failure on BAZ2A is due, at least in part, to the incorrect orientation of the structurally conserved water molecule as most crystal structures of complexes of bromodomain inhibitors provide evidence of its hydrogen-bond bridging role.[70] This interaction was unfortunately not possible in our structure. This points out again the central importance of the preparation of the system for a virtual screening campaign, to the smallest details.[75]

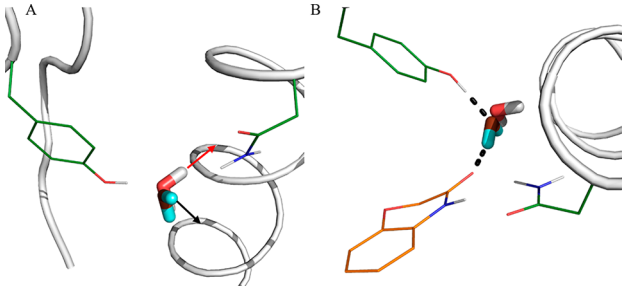


Figure 3. Orientation of the water molecule bridging to the conserved tyrosine in the structure prepared for BAZ2A (red and white sticks) vs the correct orientation as used for the other four bromodomains (purple and cyan sticks). (A) One of the two hydrogen atoms in the water of BAZ2A points toward the α-helix of the protein (red arrow) instead of pointing toward the solvent (black arrow). (B) The water molecule in the BAZ2A structure cannot act as hydrogen bond bridge with the side chain of the conserved tyrosine, a common feature found in most bromodomain inhibitors (one example is shown with carbon atoms in orange).

Two Energy Terms Used for Final Ranking Show Similar Predictive Ability. Consensus scoring has emerged as an interesting strategy to filter out false positive in virtual screening protocols.[76] Different scores are thought to compensate each other for their weaknesses. Here, we depart slightly from consensus scoring and make use of multiple ranking schemes according to individual energy contributions. Three different energy terms were used at the fragment selection stage and two of them are employed for the final ranking of the parent compounds. The three scores used, viz., Delta electrostatics, total energy efficiency, and electrostatic efficiency, favor different molecules. Delta electrostatics highlights the electrostatic complementarity between the receptor and the molecule and somewhat represents how better the molecule fits electrostatically in the protein environment rather than in water. Total energy efficiency favors molecules with a limited number of non-hydrogen atoms and is dominated by the steric complementarity between the ligand and the macromolecule (van der Waals interaction). Electrostatic efficiency emphasizes the direct electrostatic interaction between the ligand and the protein, in particular strong polar interactions such as charged contacts, weighted by the number of non-hydrogen atoms. Only Delta electrostatic and total energy efficiency were effectively used at the final compound selection step, electrostatic efficiency not selecting any unique molecule. In other words, all molecules selected because of electrostatic efficiency at the purchase stage were already, redundantly, chosen because of Delta electrostatics or total

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

energy efficiency. Of note, other combinations of scores and consensus scoring for fragments or compounds selection would also fit in the ALTA-VS strategy.

Inactive compounds (false positives in the ALTA-VS) were defined with a percentage of remnant binding between the competitor molecule and the protein of at least 65%. Binding molecules (true positives) were defined as those with an evaluated binding affinity (IC_{50} or K_D) and/or crystal structure, although not all molecules with a remnant substrate binding <65% were evaluated in dose−response assays. Of the 192 inactive compounds (i.e., false positives), 66 were chosen only because of Delta electrostatics, 60 because of total energy efficiency, and 66 were common to both energy terms used for ranking (Table 4 and Supporting Information, section S2).

Table 4. Analysis of Energy Terms Used for Ranking[a]

	inactive in single-dose measurement (N = 192)	active in single-dose measurement (N = 68)	IC_{50}, K_D, or X_{tal} (N = 27)
chosen by Delta electrostatics	132	47	17
chosen by total energy efficiency	125	36	15
chosen *only* by Delta electrostatics	66	22	10
chosen *only* by total energy efficiency	60	15	10
chosen by both Delta electrostatics and total energy efficiency	66	31	7

[a]A total of 260 molecules were tested in single-dose assays, which yielded 192 inactive compounds using a threshold of less than 65% remaining inhibition of the competitor ligand. Out of 68 molecules qualifying as active, 39 were chosen for dose−response assays, and 26 had a measurable IC_{50} (ATAD2, BRD4(1), or CREBBP) or K_D (BAZ2A or BAZ2B). In addition, compound **30** was not active in the dose−response assay (with a starting concentration of 50 μM), but its structure in complex with its target bromodomain (BAZ2B) was solved.

Thus, the three possible combinations of the two energy terms tend to produce equally false positives. Interestingly, the same is observed for true positives. Of the 27 actives (26 dose−response actives and 1 dose−response inactive but with a crystal structure), 10 (37%) were selected only according to Delta electrostatics ranking, 10 (37%) only according to total energy efficiency, and 7 (26%) because of both terms (Table 4 and Supporting Information, section S2). Thus, the two energy terms used for the final selection of the compounds show similar sensitivity and specificity.

Advantages of Fragment-Tethered High-Throughput Docking. The ALTA-VS protocol presented in this paper has several major advantages. The scoring of compounds includes solvation energies which is crucial to filter out false positives.[13] The scoring function is based on a transferable force field and thus does not require a training set for fitting. Another pivotal asset is the placement of fragment head-groups in the binding pocket. Docking of rigid fragments decreases the complexity of the problem and allows for exhaustive sampling of the positions in the binding pocket. Placing fragments in binding pockets with SEED proved to have very good predictive power.[5,7,9] Thanks to the correctly placed head-groups, the subsequent tethered docking also reduces the risks of failure by reducing the complexity of the sampling. Only conformations of the molecule around the tethered fragment need to be sampled and ranked. This step could be described as a fragment to hit expansion step. The goal is to grow from low affinity (i.e., high micromolar to millimolar) fragment binders to low micromolar hits whose potency can be interesting as starting points for further optimization in a medicinal chemistry pipeline.

Another big advantage of the ALTA-VS protocol is its cost and time efficiency. For each target, the small molecule library of nearly 77 000 compounds was investigated by docking only about 6000 fragments, followed by fragment-anchored docking of 1000 to 3000 parent molecules (more precisely 10 000 to 20 000 conformers), generating 100 000 to 200 000 poses for final minimization and finite-difference Poisson calculations (Table 1). If we consider a simple flexible docking protocol on the 77 000 molecules and extrapolate from our numbers, we would have ended up with roughly 10 million poses to be minimized per target, meaning an overall cost of ∼50 million finite-difference Poisson single point calculations for this project. On a relatively small protein like a bromodomain, the ∼50 million finite-difference Poisson calculations would have required ∼1 to 2 million CPU h on a cluster with commodity processors (Xeon E3-1280 processors at 2.5 GHz in our study). Furthermore, it is important to note that the ALTA-VS relies solely on open-source/free for not-for-profit institutions software (Table 5), namely SEED (docking and scoring engine for fragments), RDock (flexible docking engine), CHARMM (scoring engine, in particular the version

Table 5. Software Used for the ALTA-VS[a]

software	source type, licensing	ALTA-VS usage	link
RDKit	open source, Creative Commons Attribution−ShareAlike 4.0	small molecule calculations, data processing	http://www.rdkit.org/docs/Install.html
Open Babel	open source, GNU GPL	file handling	http://openbabel.org/wiki/Category:Installation
Marvin Suite (ChemAxon)	proprietary, free for noncommercial use	small molecule calculations	https://www.chemaxon.com/download/marvin-suite/
CGenFF	proprietary, permitted academic purpose	small molecule parametrization	https://www.paramchem.org/tech.php
SEED	binary executable free for all purpose	rigid docking, scoring	http://www.biochem-caflisch.uzh.ch/download/
RDock	open source, GNU LGPLv3	tethered flexible docking	http://rdock.sourceforge.net/download/
CHARMM	proprietary, free version without DOMDEC and GPU high performance modules for academic and nonprofit laboratories	minimization, scoring	http://charmm.chemistry.harvard.edu/charmm_lite.php

[a]All software are either free for academics and/or open source.

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

available at no cost for not-for-profit institutions), CGenFF (small molecule parametrizer), RDKit (data processor), Open Babel (data processor), and ChemAxon's Marvin Suite (small molecule parameters calculator).

Limitations. There are limitations in the ALTA-VS approach. Some are inherent to all virtual screening campaigns and some are more specific (although not restricted) to our (re)scoring scheme. First of all, as exemplified by the failed BAZ2A campaign, the preparation of the input structures (protein target, compound library) can strongly influence the results. Furthermore, in most human bromodomains the binding pocket for the natural ligand is rather small and rigid and can be defined by one or two conserved residues, which is not necessarily the case for other targets, e.g., proteases. For those proteins with large and/or flexible binding site, the SEED program could be used in a preprocessing step with a small subset of fragments to target a broadly defined binding site and/or multiple conformers originating from molecular dynamics[77] or crystal structures.[8] In this way, individual subpockets with favorable binding energy for the probe fragments could be identified from multiple structures. More specifically to our protocol, the ALTA-VS is based on single-point energy calculations; entropy and the contribution of different metastable states (e.g., multiple orientations of the nonanchor moiety) to the free energy of binding are neglected. Additionally, the scoring relies on the continuum dielectric approximation in both the fragment and full molecule stages. Thus, it suffers from the theoretical limitations of such models, like the depiction of the protein as a macroscopic dielectric medium, the definition of the dielectric boundary, and the disregard of nonpolar contributions to the hydration free energy.[78] However, the assumptions made seem to be a good compromise in terms of computational performance versus quality of the scoring, when having to rescore hundreds of thousands or even millions of poses.[44] Moreover, the quality of the scoring calculations also depends on the quality of the force field used. We used state of the art and fully consistent force fields CHARMM36 for the protein[15] and CHARMM generalized force field (CGenFF)[16,17] for all the fragments and parent molecules of the library used for ALTA-VS.

■ CONCLUSIONS

The number of commercially available small organic molecules is growing steadily and is already close to one hundred million.[79] Such continuous growth calls for efficient protocols for in silico screening, particularly for protein structure-based methods. The ALTA-VS protocol for fragment-anchored ligand docking (Figure 1), introduced in 2005 and further improved for the present application to bromodomains, combines the advantages of docking with those of fragment-based approaches. Four main results emerge from the present study. First, the updated ALTA-VS protocol is able to screen a library of nearly 77 000 compounds (about 150 000 poses for energy minimization, Table 1) within 2 days of computational time on a 96-core compute cluster (equipped with Xeon E3-1280 quad core processors at 2.5 GHz). Most of the computational time is required for energy minimization of the docked poses and evaluation of electrostatic solvation by the finite-difference Poisson equation which require around 1.5 min per pose. Overall, the ALTA-VS has identified ligands for four bromodomains (Table 2). More precisely, 15 of the 39 compounds for which dose−response measurements were performed have an affinity of 100 μM or better (Table 3).

Second, the application of the updated ALTA-VS protocol provides evidence of the usefulness of the novel aspects. Retrospectively, the use of a single anchor fragment in the updated ALTA-VS protocol has resulted in 15 active molecules (out of 27) consisting of only two ring systems. These ligands would not have been identified by the original version of ALTA-VS, which required three anchor fragments. Furthermore, most of these ligands have only one ring system (the headgroup) fully buried in the bromodomain binding site while the tail groups are rather flexible and can assume multiple orientations which is not congruent with the role of anchor fragment. Another new aspect of the updated ALTA-VS protocol is the use of a transferable force field without any fitting parameter. The paucity of known inhibitors of the ATAD2 bromodomain would have hindered the derivation of the fitting parameters which was necessary in the original ALTA-VS protocol.

Third, the preparation of the protein structure used for docking is a crucial step. In the ALTA-VS campaign for BAZ2A, the incorrect orientation of a structurally conserved water molecule (Figure 3) resulted in a more poor outcome than for the other four bromodomains for which the orientation of the conserved water was correct. This water molecule acts as hydrogen bond-bridge between a conserved tyrosine side chain and the natural ligand (acetylated lysine) or the large majority of known inhibitors. The correct orientation of the water hydrogens plays a critical role because the electrostatic energy, which is based on partial charges, is very sensitive on the position and orientations of dipoles, and the water molecule has a strong dipole moment of 2.35 D in the force field used in this work.

Fourth, two different energy terms were used for the final selection of compounds: Delta electrostatics (i.e., the protein/ligand electrostatic interaction minus the free energy of hydration) and total binding energy divided by the number of non-hydrogen atoms. These terms showed similar predictive ability (Table 4). More precisely, selection of compounds for in vitro testing by either of the two terms (or both) resulted in a similar amount of true positives (sensitivity) and similar numbers of false positives (specificity).

In conclusion, the present study and our previous applications of the ALTA-VS protocol[10−14] provide strong evidence of the usefulness and efficiency of fragment-anchored docking of flexible molecules for in silico screening. The ALTA-VS protocol can be employed also for screening for covalent binders by replacing tethered docking with covalent docking. If experimental binding modes of fragment hits are available, they can be integrated in the protocol and replace docked pose for tethered docking. Since the updated version of the ALTA-VS makes use of a transferable force field and does not require any training set for fitting parameters, it should be applicable to most target proteins of known three-dimensional structure. It also has the notable advantage to be based solely on open source and/or free for academics software.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.7b00336.

Detailed description of the screening library, details of docking scores with respect to experimental data, Alphascreen/BROMOscan dose−response curves, X-ray

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

crystallization data, docking and experimental data for all tested molecules (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62. E-mail: caflisch@bioc.uzh.ch.

ORCID

Jean-Rémy Marchand: 0000-0002-8002-9457

Amedeo Caflisch: 0000-0002-2317-6792

Author Contributions

The study was designed by J.R.M. and A.C. J.R.M. performed the docking. J.R.M. and A.C. analyzed the docking results. A.D.V. and G.L. carried out the crystallography. The manuscript was written by J.R.M. and A.C. All authors have given approval to the final version of the manuscript.

Funding

Swiss Cancer Society (Krebsliga Schweiz, grant Nr. KLS-3098-02-2013).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Carmen Esposito, Nicholas Deerain, and Dr. Dimitrios Spiliotopoulos for compound ordering and handling. The authors also thank the anonymous reviewers for their valuable comments that helped improve the clarity and scope of the manuscript.

■ ABBREVIATIONS

ALTA-VS, anchor-based library tailoring approach for virtual screening; ATAD2, ATPase family AAA domain containing 2; BAZ2A, bromodomain adjacent to zinc finger domain 2A; BAZ2B, bromodomain adjacent to zinc finger domain 2B; BRD4(1), bromodomain-containing protein 4, bromodomain 1; CREBBP, cAMP response element binding protein binding protein; DMSO, dimethyl sulfoxide; IMAC, immobilized metal ion affinity chromatography; MPD, 2-methyl-2,4-pentanediol; TEV protease, tobacco etch virus nuclear-inclusion-a endopeptidase

■ REFERENCES

(1) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discovery* **2016**, *15*, 605−619.

(2) Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59*, 4103−4120.

(3) Seifert, M. H. J.; Wolf, K.; Vitt, D. Virtual High-Throughput In Silico Screening. *BIOSILICO* **2003**, *1*, 143−149.

(4) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213−2221.

(5) Spiliotopoulos, D.; Zhu, J.; Wamhoff, E. C.; Deerain, N.; Marchand, J. R.; Aretz, J.; Rademacher, C.; Caflisch, A. Virtual Screen to NMR (VS2NMR): Discovery of Fragment Hits for the CBP Bromodomain. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 2472−2478.

(6) Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **2015**, *55*, 2411−2420.

(7) Lolli, G.; Caflisch, A. High-Throughput Fragment Docking into the BAZ2B Bromodomain: Efficient in Silico Screening for X-Ray Crystallography. *ACS Chem. Biol.* **2016**, *11*, 800−807.

(8) Xu, M.; Unzue, A.; Dong, J.; Spiliotopoulos, D.; Nevado, C.; Caflisch, A. Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics. *J. Med. Chem.* **2016**, *59*, 1340−1349.

(9) Zhu, J.; Caflisch, A. Twenty Crystal Structures of Bromodomain and PHD Finger Containing Protein 1 (BRPF1)/Ligand Complexes Reveal Conserved Binding Motifs and Rare Interactions. *J. Med. Chem.* **2016**, *59*, 5555−5561.

(10) Huang, D.; Luthi, U.; Kolb, P.; Cecchini, M.; Barberis, A.; Caflisch, A. In Silico Discovery of Beta-Secretase Inhibitors. *J. Am. Chem. Soc.* **2006**, *128*, 5436−5443.

(11) Huang, D.; Luthi, U.; Kolb, P.; Edler, K.; Cecchini, M.; Audetat, S.; Barberis, A.; Caflisch, A. Discovery of Cell-Permeable non-Peptide Inhibitors of Beta-Secretase by High-Throughput Docking and Continuum Electrostatics Calculations. *J. Med. Chem.* **2005**, *48*, 5108−5111.

(12) Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-Based Tailoring of Compound Libraries for High-Throughput Screening: Discovery of Novel EphB4 Kinase Inhibitors. *Proteins: Struct., Funct., Genet.* **2008**, *73*, 11−18.

(13) Kolb, P.; Huang, D.; Dey, F.; Caflisch, A. Discovery of Kinase Inhibitors by High-Throughput Docking and Scoring Based on a Transferable Linear Interaction Energy Model. *J. Med. Chem.* **2008**, *51*, 1179−1188.

(14) Schenker, P.; Alfarano, P.; Kolb, P.; Caflisch, A.; Baici, A. A Double-Headed Cathepsin B Inhibitor Devoid of Warhead. *Protein Sci.* **2008**, *17*, 2145−2155.

(15) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(16) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.

(17) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(18) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012**, *149*, 214−231.

(19) Marchand, J. R.; Caflisch, A. Binding Mode of Acetylated Histones to Bromodomains: Variations on a Common Motif. *ChemMedChem* **2015**, *10*, 1327−1333.

(20) Fujisawa, T.; Filippakopoulos, P. Functions of Bromodomain-Containing Proteins and their Roles in Homeostasis and Cancer. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 246−262.

(21) United States National Library of Medicine. www.clinicaltrials.gov.

(22) Theodoulou, N. H.; Tomkinson, N. C. O.; Prinjha, R. K.; Humphreys, P. G. Clinical Progress and Pharmacology of Small Molecule Bromodomain Inhibitors. *Curr. Opin. Chem. Biol.* **2016**, *33*, 58−66.

(23) Muller, S.; Filippakopoulos, P.; Knapp, S. Bromodomains as Therapeutic Targets. *Expert Rev. Mol. Med.* **2011**, *13*, e29.

(24) Unzue, A.; Lafleur, K.; Zhao, H.; Zhou, T.; Dong, J.; Kolb, P.; Liebl, J.; Zahler, S.; Caflisch, A.; Nevado, C. Three Stories on Eph Kinase Inhibitors: From In Silico Discovery to In Vivo Validation. *Eur. J. Med. Chem.* **2016**, *112*, 347−366.

(25) Zhao, H.; Caflisch, A. Molecular Dynamics in Drug Design. *Eur. J. Med. Chem.* **2015**, *91*, 4−14.

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(26) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive Docking of Molecular Fragments with Electrostatic Solvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88−105.
(27) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient Electrostatic Solvation Model for Protein-Fragment Docking. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 256−268.
(28) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.
(29) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10*, 10.
(30) Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A. Automated Docking of Highly Flexible Ligands by Genetic Algorithms: A Critical Assessment. *J. Comput. Chem.* **2004**, *25*, 412−422.
(31) Budin, N.; Majeux, N.; Caflisch, A. Fragment-Based Flexible Ligand Docking by Evolutionary Optimization. *Biol. Chem.* **2001**, *382*, 1365−1372.
(32) Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* **2012**, *55*, 7346−7359.
(33) Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: Computation of Electrostatic Forces from Numerical Solutions to the Poisson-Boltzmann Equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(34) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(35) EPFL Biomolecular Screening Facility, BSF-ACCESS Chemicals Collections. http://bsf.epfl.ch/collections#faq-568582.
(36) ChemAxon Prediction of dissociation constant using micro-constants. https://docs.chemaxon.com/download/attachments/41128347/Prediction_of_dissociation_constant_using_microconstants.pdf.
(37) Landrum, G. RDKit: Open-source cheminformatics. rdkit.org.
(38) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.
(39) Filippakopoulos, P.; Knapp, S. Targeting Bromodomains: Epigenetic Readers of Lysine Acetylation. *Nat. Rev. Drug Discovery* **2014**, *13*, 337−356.
(40) Ferguson, F. M.; Fedorov, O.; Chaikuad, A.; Philpott, M.; Muniz, J. R.; Felletar, I.; von Delft, F.; Heightman, T.; Knapp, S.; Abell, C.; Ciulli, A. Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain. *J. Med. Chem.* **2013**, *56*, 10183−10187.
(41) Marchand, J. R.; Lolli, G.; Caflisch, A. Derivatives of 3-Amino-2-Methylpyridine as BAZ2B Bromodomain Ligands: In Silico Discovery and in Crystallo Validation. *J. Med. Chem.* **2016**, *59*, 9919−9927.
(42) Chaikuad, A.; Petros, A. M.; Fedorov, O.; Xu, J.; Knapp, S. Structure-Based Approaches towards Identification of Fragments for the Low-Druggability ATAD2 Bromodomain. *MedChemComm* **2014**, *5*, 1843−1848.
(43) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using what we Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.
(44) Foloppe, N.; Hubbard, R. Towards Predictive Ligand Design with Free-Energy Based Computational Methods? *Curr. Med. Chem.* **2006**, *13*, 3583−3608.
(45) Hunenberger, P. H.; Helms, V.; Narayana, N.; Taylor, S. S.; McCammon, J. A. Determinants of Ligand Binding to cAMP-Dependent Protein Kinase. *Biochemistry* **1999**, *38*, 2358−2366.
(46) Froloff, N.; Windemuth, A.; Honig, B. On the Calculation of Binding Free Energies Using Continuum Methods: Application to MHC Class I Protein-Peptide Interactions. *Protein Sci.* **1997**, *6*, 1293−1301.
(47) Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and Free Energy Analyses of Cathepsin D-Inhibitor Interactions: Insight into Structure-Based Ligand Design. *J. Med. Chem.* **2002**, *45*, 1412−1419.
(48) Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S. Can the Calculation of Ligand Binding Free Energies be Improved with Continuum Solvent Electrostatics and an Ideal-Gas Entropy Correction? *J. Comput. Chem.* **2002**, *23*, 1143−1149.
(49) Shen, J.; Wendoloski, J. Electrostatic Binding Energy Calculation Using the Finite Difference Solution to the Linearized Poisson-Boltzmann Equation: Assessment of its Accuracy. *J. Comput. Chem.* **1996**, *17*, 350−357.
(50) Czaplewski, C.; Ripoll, D. R.; Liwo, A.; Rodziewicz-Motowidlo, S.; Wawak, R. J.; Scheraga, H. A. Can Cooperativity in Hydrophobic Association Be Reproduced Correctly by Implicit Solvation Models? *Int. J. Quantum Chem.* **2002**, *88*, 41−55.
(51) Feig, M.; Brooks, C. L. Recent Advances in the Development and Application of Implicit Solvent Models in Biomolecule Simulations. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.
(52) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523−9530.
(53) Pitera, J. W.; van Gunsteren, W. F. The Importance of Solute-Solvent Van Der Waals Interactions with Interior Atoms of Biopolymers. *J. Am. Chem. Soc.* **2001**, *123*, 3163−3164.
(54) Shimizu, S.; Chan, H. S. Anti-Cooperativity and Cooperativity in Hydrophobic Interactions: Three-Body Free Energy Landscapes and Comparison with Implicit-Solvent Potential Functions for Proteins. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 15−30.
(55) Philpott, M.; Yang, J.; Tumber, T.; Fedorov, O.; Uttarkar, S.; Filippakopoulos, P.; Picaud, S.; Keates, T.; Felletar, I.; Ciulli, A.; Knapp, S.; Heightman, T. D. Bromodomain-Peptide Displacement Assays for Interactome Mapping and Inhibitor Discovery. *Mol. BioSyst.* **2011**, *7*, 2899−2908.
(56) Quinn, E.; Wodicka, L.; Ciceri, P.; Pallares, G.; Pickle, E.; Torrey, A.; Floyd, M.; Hunt, J.; Treiber, D. Abstract 4238: BROMOscan - a High Throughput, Quantitative Ligand Binding Platform Identifies Best-in-Class Bromodomain Inhibitors from a Screen of Mature Compounds Targeting other Protein Classes. *Cancer Res.* **2013**, *73*, 4238.
(57) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E. Selective Inhibition of BET Bromodomains. *Nature* **2010**, *468*, 1067−1073.
(58) Hammitzsch, A.; Tallant, C.; Fedorov, O.; O'Mahony, A.; Brennan, P. E.; Hay, D. A.; Martinez, F. O.; Al-Mossawi, M. H.; de Wit, J.; Vecellio, M.; Wells, C.; Wordsworth, P.; Muller, S.; Knapp, S.; Bowness, P. CBP30, a Selective CBP/p300 Bromodomain Inhibitor, Suppresses Human Th17 Responses. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 10768−10773.
(59) 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−W207.
(60) Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K. An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **2015**, *58*, 7076−7087.
(61) Spiliotopoulos, D.; Wamhoff, E.-C.; Lolli, G.; Rademacher, C.; Caflisch, A. Discovery of BAZ2A Bromodomain Ligands. *Eur. J. Med. Chem.* **2017**, DOI: 10.1016/j.ejmech.2017.08.028.

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(62) Lolli, G.; Battistutta, R. Different Orientations of Low-Molecular-Weight Fragments in the Binding Pocket of a BRD4 Bromodomain. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 2161−2164.
(63) Kabsch, W. Xds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(64) Evans, P. R.; Murshudov, G. N. How Good Are my Data and What Is the Resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−1214.
(65) Karplus, P. A.; Diederichs, K. Assessing and Maximizing Data Quality in Macromolecular Crystallography. *Curr. Opin. Struct. Biol.* **2015**, *34*, 60−68.
(66) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(67) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.
(68) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(69) Bamborough, P.; Chung, C. W.; Demont, E. H.; Furze, R. C.; Bannister, A. J.; Che, K. H.; Diallo, H.; Douault, C.; Grandi, P.; Kouzarides, T.; Michon, A. M.; Mitchell, D. J.; Prinjha, R. K.; Rau, C.; Robson, S.; Sheppard, R. J.; Upton, R.; Watson, R. J. A Chemical Probe for the ATAD2 Bromodomain. *Angew. Chem., Int. Ed.* **2016**, *55*, 11382−11386.
(70) Crawford, T. D.; Tsui, V.; Flynn, E. M.; Wang, S.; Taylor, A. M.; Cote, A.; Audia, J. E.; Beresini, M. H.; Burdick, D. J.; Cummings, R.; Dakin, L. A.; Duplessis, M.; Good, A. C.; Hewitt, M. C.; Huang, H. R.; Jayaram, H.; Kiefer, J. R.; Jiang, Y.; Murray, J.; Nasveschuk, C. G.; Pardo, E.; Poy, F.; Romero, F. A.; Tang, Y.; Wang, J.; Xu, Z.; Zawadzke, L. E.; Zhu, X.; Albrecht, B. K.; Magnuson, S. R.; Bellon, S.; Cochran, A. G. Diving into the Water: Inducible Binding Conformations for BRD4, TAF1(2), BRD9, and CECR2 Bromodomains. *J. Med. Chem.* **2016**, *59*, 5391−5402.
(71) Tanaka, M.; Roberts, J. M.; Seo, H. S.; Souza, A.; Paulk, J.; Scott, T. G.; DeAngelo, S. L.; Dhe-Paganon, S.; Bradner, J. E. Design and Characterization of Bivalent BET Inhibitors. *Nat. Chem. Biol.* **2016**, *12*, 1089−1096.
(72) Theodoulou, N. H.; Bamborough, P.; Bannister, A. J.; Becher, I.; Bit, R. A.; Che, K. H.; Chung, C. W.; Dittmann, A.; Drewes, G.; Drewry, D. H.; Gordon, L.; Grandi, P.; Leveridge, M.; Lindon, M.; Michon, A. M.; Molnar, J.; Robson, S. C.; Tomkinson, N. C. O.; Kouzarides, T.; Prinjha, R. K.; Humphreys, P. G. Discovery of I-BRD9, a Selective Cell Active Chemical Probe for Bromodomain Containing Protein 9 Inhibition. *J. Med. Chem.* **2016**, *59*, 1425−1439.
(73) Flynn, E. M.; Huang, O. W.; Poy, F.; Oppikofer, M.; Bellon, S. F.; Tang, Y.; Cochran, A. G. A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. *Structure* **2015**, *23*, 1801−1814.
(74) Unzue, A.; Zhao, H.; Lolli, G.; Dong, J.; Zhu, J.; Zechner, M.; Dolbois, A.; Caflisch, A.; Nevado, C. The "Gatekeeper" Residue Influences the Mode of Binding of Acetyl Indoles to Bromodomains. *J. Med. Chem.* **2016**, *59*, 3087−3097.
(75) Madhavi Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(76) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−1426.
(77) Ekonomiuk, D.; Su, X. C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Otting, G.; Huang, D.; Caflisch, A. Flaviviral Protease Inhibitors Identified by Fragment-Based Library Docking into a Structure Generated by Molecular Dynamics. *J. Med. Chem.* **2009**, *52*, 4860−4868.
(78) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(79) Sterling, T.; Irwin, J. J. ZINC 15-Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.

DOI: 10.1021/acs.jcim.7b00336
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX