Fragment-based in silico screening of bromodomain ligands

Dimitrios Spiliotopoulos*, Amedeo Caflisch*

Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

We review the results of fragment-based high-throughput docking to the N-terminal bromodomain of BRD4 and the CREBBP bromodomain. In both docking campaigns the ALTA (anchor-based library tailoring) procedure was used to reduce the size of the initial library by selecting for flexible docking only the molecules that contain a fragment with favorable predicted binding energy. Ranking by a force field-based energy with solvation has resulted in small-molecule hits with low-micromolar affinity and favorable ligand efficiency. Importantly, the binding modes predicted by docking have been validated by X-ray crystallography. One of the hits for the CREBBP bromodomain has been optimized by medicinal chemistry into a series of potent and selective ligands.

Introduction

Fragment-based drug design is an efficient and cost-effective alternative to high throughput screening [1]. The main reason for screening fragments rather than compounds is valid for both in vitro and in silico approaches; a collection of 10^4 diverse fragments (MW < 250 Da) usually has a larger chemical diversity than a library of 10^6–10^7 molecules (MW < 500 Da). Furthermore, fragments tend to have good solubility which is advantageous for in vitro biophysical assays. An important advantage for in silico screening is that the docking of mainly rigid fragments is always more efficient and usually more accurate than docking molecules with several rotatable bonds [2]. Fragment-based screening campaigns have been reported as the starting point for the development of selective and cell-active bromodomain ligands both in vitro [3–20] and in silico (vide infra).

Our group has developed a fragment-based procedure for high-throughput docking of large libraries of compounds. The procedure, called ALTA (anchor-based library tailoring), was originally published in 2008 with an application to the EphB4 tyrosine kinase [21], and applications to other protein targets have been reviewed recently [22]. ALTA starts by the decomposition of the compound library into rigid fragments followed by docking and ranking of the fragments. The top-ranking fragments (called also anchor fragments) are used as query to retrieve the compounds they originate from. The ‘parent’ compounds are usually a relatively small subset of the original library so that their docking requires a fraction of the computational cost. An additional benefit of the ALTA procedure is that the top-ranking fragments that are commercially available can be purchased for soaking into the apo crystals of the protein target.

Here, we first review in silico screening campaigns by others, and then present the applications of our ALTA procedure to two human bromodomains belonging to different subfamilies: the N-terminal bromodomain of the bromodomain-containing protein 4 (BRD4(1)) and the bromodomain

*Corresponding authors: D. Spiliotopoulos (d.spiliotopoulos@bioc.uzh.ch), A. Caflisch (caflisch@bioc.uzh.ch)

Section editor:
Ming-Ming Zhou – Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, United States.
of the binding protein of the cAMP-response element binding protein (CREB)-binding protein (CREBBP). In both cases, low-micromolar ligands (Fig. 1) were identified by the ALTA procedure in less than two weeks on 100 cores of a compute cluster of commodity processors.

In silico screening for human bromodomains
Bromodomains are epigenetic reader modules that bind acetylated lysine and have become the object of intense investigations aimed at unravelling their biological function and targeting them with small molecules [23,24]. Recent reports have compiled the structural details of their interaction with posttranslationally modified peptides [25,26] and small molecule inhibitors [27]. In silico approaches have been undertaken as the starting point of the development of selective, cell-active inhibitors targeting bromodomains. Three papers focused on BRD4(1). A high-throughput virtual screening was performed on BRD4(1) selecting 22 small molecules out of the original 7 millions, with the best compound having a $K_{d}$ of 33 µM [28]. The Structural Genomics Consortium purchased 250 compounds, with two of them yielding micromolar $IC_{50}$ values (4.7 µM and 80.9 µM) as measured by the AlphaScreen assay [29]. More recently, Xue reported on the screening of 15 compounds from an in-house library of 10,000, with the best performing compounds having single-digit micromolar $IC_{50}$ values when tested via AlphaScreen [30]. Few reports have appeared on the use of computational methods to identify fragment hits for non-BET bromodomains. Based on previous experimental indications [4,31,32], ~6000 benzoxazinone amines were docked into the CREBBP bromodomain (with the best performing small molecules having $IC_{50}$ values of 51 µM and 77 µM via AlphaScreen): inhibitors have subsequently been developed expanding a dihydroquinoxalinone series [33]. A report on the development of BRD9 inhibitors relied on the structural information of the BRD4(1) bromodomain in the complex with a pyrimidine ligand and focused

Figure 1. Flowcharts of the applications of the ALTA procedure for in silico screening. The steps are numbered as follows: (1) compound decomposition and fragment filtering; (2) rigid fragment docking and filtering; (3) retrieval of the parent compounds for the most favorable fragments; (4) flexible compound docking and filtering; (5) actives on purchased compounds. The X-ray crystal structure in the complex with the inhibitor JQ1 was used for BRD4(1) (PDB code 3MXF) [41]. Two crystal structures were used for ALTA screening in CREBBP, the complex with a 3,5-dimethylisoxazole ligand (3SVH) and acetylated lysine (3P1C), respectively [48,49].
its docking procedures on purine fragments [34]. Alternatively, a combination of computational and biophysical techniques were used in parallel to identify chemical probes targeting the BRD9 [35] and TRIM24-BRPF1 [20] bromodomains. There are also reports on bromodomain ligands identified using a computational approach but no information is given on their selectivity and/or activity on cells [36–40].

**Discovery of BRD4(1) bromodomain inhibitors**

This subsection reviews the in silico screening campaign reported in [41] (Fig. 1, left). First the nearly nine million compounds in the 2012-version of the ZINC all-now library [42] were reduced to about half by four criteria related to the 2D structure: molecular weight between 200 and 400 Da, two or more hydrogen bond acceptors, less than eight rotatable bonds, and two to six rings. The remaining 4.6 million compounds were decomposed automatically into 375,897 fragments by an algorithm that (1) breaks all rotatable bonds to obtain initial ring systems, each of which (2) is next extended until it reaches another ring or the number of heavy atoms of the extension reaches three. The algorithm then (3) adds hydrogen atoms to maintain the original valence and (4) parses again the list of fragments to identify double entries and retain only unique molecules [43]. These fragments were then filtered by molecular weight between 60 Da and 300 Da, at least one hydrogen bond acceptor, and up to three rotatable bonds. The remaining 238,408 fragments were docked into the structure of BRD4(1) (in the complex with the small-molecule inhibitor JQ1, PDB code 3MXF) by a genetic algorithm-based docking program [41,44] which generated 511,417 poses. The poses obtained by docking were then minimized in the rigid protein by CHARMM [45]. Two additional filters were applied after minimization: the presence of a hydrogen bond to the conserved asparagine (Asn140 in BRD4(1)), and a value of the hydrogen bonding penalty smaller than 1. The latter filters removes poses with buried polar groups of ligand and/or protein not involved in hydrogen bonds [46]. For each fragment the binding energy was calculated by a force field-based energy function which takes into account solvation [43]. Overall, the docking of the 238,408 fragments required less than one day of a compute cluster of 100 cores of Xeon 2.6 GHz processors while the CHARMM minimization of 511,417 poses followed by scoring with solvation took 2.7 days on the same cluster. Two filters were applied for the selection of anchor fragments: a predicted binding affinity more favorable than 4 kcal/mol (which corresponds to 1 mM at room temperature) and a predicted ligand efficiency more favorable (i.e., larger) than 0.4 kcal/mol per non-hydrogen atom.

In the second step of the ALTA procedure, the 665,184 molecules containing at least one of the 17,179 anchor fragments were submitted to flexible ligand docking followed by minimization which required nearly five days on the same 100-core cluster. The final filtering was based again on predicted binding energy and ligand efficiency with thresholds of –6.5 kcal/mol and 0.3 kcal/mol per heavy atom, respectively. The remaining compounds (near- ly 5000 molecules for a total of about 1000 anchor fragments) were clustered using ECFP4 fingerprints [47] which resulted in 616 cluster. A subset of 55 cluster representatives was selected for explicit solvent molecular dynamics. This selection was based on favorable ligand efficiency, chemical diversity, rigidity, and novelty. Two 100-ns molecular dynamics simulations were carried out for each ligand to analyze the structural (i.e., kinetic) stability of the predicted binding mode with particular emphasis on the main interactions, for example, the hydrogen bond with the side chain of the conserved Asn140. These simulations revealed that some of the chemotypes were not stable in their predicted binding mode (i.e., rupture of the hydrogen bond to the side chain of Asn140 was observed), and these molecules were filtered out.

Finally, only 24 compounds were purchased for experimental validation and four of them (Fig. 2a) showed activity at 50 μM in an AlphaScreen assay (performed at Reaction Biology Corp, Malvern, PA). The value of the inhibitor concentration for a 50% response (IC50) was 7.0 μM and 7.5 μM for compounds 2 and 3, respectively, so that their ligand efficiency is very favorable (0.37 kcal/mol per non-hydrogen atom). Importantly, for both compounds the binding mode predicted by docking is essentially identical as the one observed in the crystal structures (PDB codes 4PCE and 4PCI, Fig. 2b,c, respectively). The carbonyl group of compounds 2 and 3 is a hydrogen bond acceptor for the side chain of the conserved Asn140 and the structural water molecule that acts as bridge to the conserved Tyr97. For both compounds, the hydrogen bond with the Tyr97-bridging water molecule has a shorter distance (2.8 Å) than the hydrogen bond with the conserved Asn140 (3.0 Å). While the carbonyl groups of compounds 2 and 3 occupy the same position and are involved in the same interactions with the BRD4(1) bromodomain, the phenyl rings point in opposing directions. The phenyl of compound 2 is involved in hydrophobic contacts with the so-called gatekeeper residue (Ile146) and Trp81 (which is the first residue of the WPF triad) while the phenyl of compound 3 is oriented towards the side chains of Leu92 and Leu94 which are both in the ZA-loop (i.e., the loop that bridges the helices Z and A). It is interesting to note that only one of the two enantiomers of ligand 3 was predicted by docking to bind which was confirmed by the X-ray structure. In conclusion, two novel chemotypes of inhibitors of the BRD4(1) bromodomain have been identified by fragment-based high-throughput docking followed by pose minimization and molecular dynamics simulations of the top ranking molecules.
Discovery of CREBBP bromodomain inhibitors

This subsection reviews the in silico screening campaign reported in [48,49] (Fig. 1, right). The crystal structures of the CREBBP bromodomain available when this campaign started were visually inspected to assess the flexibility of the residues forming the binding pocket. Two residues captured our attention, namely the so-called ‘gatekeeper’ residue, Val1174 [50], and a solvent-exposed residue, Arg1173, that had been suggested to be implicated in the binding selectivity towards the CREBBP bromodomain with respect to other human bromodomains for small molecules [5,33,51] and acetyl lysine peptides [52]. To account, at least partially, for the flexibility of these side chains, we decided to utilize two structures for the ALTA procedure (Fig. 1, right), that is, the CREBBP bromodomain bound to acetyl-lysine and a 3,5-dimethylisoxazole ligand (PDB codes 3P1C and 3SVH, respectively) [48].

The October 2012 version of the ZINC leads-now library, comprising nearly 2 million compounds, was decomposed using the DAIM program [53]. Only the fragments containing a hydrogen bond donor or acceptor and a ring were kept and parametrized using MATCH [54]. The remaining ~97,000 fragments were docked into the two CREBBP bromodomain structures using SEED [55,56], a software designed to dock rigid fragments. The docked fragments were reduced to approximately 4,000 per structures by applying two filters: (a) a predicted ligand efficiency more favorable than 0.125 kcal/mol per gram and (b) a hydrogen bonding penalty no greater than 1. Of note, these filters increased the percentage of fragments establishing a hydrogen bond with the conserved asparagine side chain (Asn1168, in the CREBBP bromodomain). The ALTA procedure then progressed as the parent compounds of the best ranking fragments (Table 1) were retrieved, docked using AutoDock Vina [57] (which can perform flexible docking) and the poses were minimized with CHARMM [45]. The lowest-energy poses were rescored according to the affinity estimated via a force field-based energy function provided by SEED [55,56]. This energy function takes into account van der Waals and electrostatic terms, including both intermolecular electrostatics and desolvation energies. Remarkably, only twenty molecules were sufficient to describe the 1000 top-ranking compounds in terms of

Figure 2. BRD4(1) bromodomain hits identified by the ALTA procedure and validation of binding mode by X-ray crystallography. (a) Chemical structures of the four BRD4(1) bromodomain hits [41]. The stereogenic center of compound 3 is highlighted with an asterisk. Compound 4 is diazepam, a benzodiazepine first marketed as Valium, which is one of the most prescribed drug since its launch in 1963. Values in percentage indicate residual binding of the lysine-acetylated peptide, thus lower percentages indicate stronger binding of the compound. The IC50 value is the compound concentration that inhibits binding to the lysine-acetylated peptide by 50%. (b,c) Crystal structures of the complexes of the BRD4(1) bromodomain (cyan ribbon) and the hits 2 and 3 (carbon atoms in yellow) identified by the ALTA procedure (PDB codes 4PCE and 4PCI, respectively). The hydrogen bonds mentioned in the text are highlighted (black dashed lines).
chemical space of the functional groups that interact with the conserved asparagine (called head groups in the following).

To assess the stability of these 20 compounds in the CREBBP bromodomain binding pocket, a 100-ns molecular dynamics simulation was performed starting from each docked pose. For five of the 20 compounds an alternative binding mode was revealed by the simulations. Three of the small molecules moved out of the binding site (heavy-atom RMSD from the docked pose or alternative pose larger than 3Å), and thus only 17 compounds were purchased to be experimentally validated via a competition binding assay (BROMOScan, performed at DiscoveRx) [58,59]. Gratifyingly, five compounds (compounds 5–9, Fig. 3a) showed activity at 50 μM. The values of the equilibrium dissociation constant (K_D) of compounds 8 and 9 are 13 μM and 29 μM, respectively, corresponding to a ligand efficiency of 0.27 kcal/mol per non-hydrogen atom [48]. The promising properties of these compounds bearing acylaryl head groups (viz., acetylpyrrole and acetylbenzene) prompted us to engage in medicinal chemistry efforts aimed to optimize their affinity and selectivity [49].

At the beginning of the optimization of the acetylpyrrole compound 8, another research group identified acetylpyrrole

<table>
<thead>
<tr>
<th>Anchor</th>
<th>ChEpiMod</th>
<th>PDBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>19</td>
<td>–</td>
</tr>
<tr>
<td>F2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>F3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F4</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>F5</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>F6</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>F7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>F10</td>
<td>45</td>
<td>23</td>
</tr>
</tbody>
</table>

*The most frequent anchor fragments in the 1000 top ranked compounds of the high-throughput docking campaign targeting the CREBBP bromodomain [48]. The methyl groups in green are H atoms in some of the compounds.

*The number of 'Hits’, ‘Active’, and ‘Inconclusive’ for CREBBP are reported (searches are updated to May 2016).

*K_D and IC_{50} values are given in μM.

*Thermal shifts values (ΔT_m) are given in °C. Several bromodomain ligands, including I-BET151 [66,67] and SGC bromodomain inhibitors [8] present a 3,5-dimethyl-1,2-oxazole (F1) moiety. Eight of the 45 hits of fragment F10 targeted the CREBBP histone acetyltransferase domain [68].
derivatives targeting the BET bromodomains with some cross-reactivity for the CREBBP bromodomain but less favorable binding efficiency than the acetylpyrrole \( 8 \) [28]. This report and the lack of selectivity limited the interest in the identified acetylpyrrole \( 8 \), for which just two derivatives were designed based on the analysis of two independent 1-\( \mu \)s runs of explicit solvent molecular dynamics (Figure 7 of [48]). One of these two derivatives (compound 6 in Table 1 of [48]).
features a benzoate instead of the p-methyl-(1,3,4-oxadiazol-2-yl)phenyl moiety of hit 8, and shows a Kd value of 4.2 μM and a ligand efficiency of 0.34 kcal/mol per heavy atom which is a significant improvement with respect to the original hit.

We decided to focus on the optimization of the acetylbenzene hit 9 by medicinal chemistry because of its potency and novelty. This optimization campaign aimed at preserving the acetylbenzene head group while trying to improve the interaction with the Arg1173 guanidinium. Nanomolar ligands with remarkable ligand efficiency and excellent selectivity towards other human bromodomains (UZH-1 and UZH-2 in Fig. 4) were obtained by the synthesis of only twenty compounds [49]. The binding mode in the crystal structure of the complex of the CREBBP bromodomain with one of the synthesized compounds, the acetylbenzene derivative 10 (Kd value of 0.77 μM and ligand efficiency of 0.35 kcal/mol per non-hydrogen atom), confirmed the pattern of contacts predicted by docking (PDB code 4TQN, Fig. 3B). In particular, the acetyl oxygen of compound 10 acts as acceptor for the side chain of the conserved Asn1168 (oxygen to nitrogen distance of 3.0 Å) and a water-mediated interaction with the hydroxyl of Tyr1125 (with a distance of 2.8 Å between acetyl oxygen and water oxygen). Further water-mediated hydrogen bonds involve the amide of the linker, whose amine and carbonyl groups interact with the backbone oxygen of Pro1110 of the LPF triad (corresponding to the BRD4(1) WPF shelf) and the Arg1173 side chain, respectively. Finally, the carboxylic acid of 10 interacts with the Arg1173 guanidinium which, as mentioned above, is important for CREBBP bromodomain ligand selectivity [5]. According to the structure-based sequence alignment reported in [60], only four human bromodomains (CREBBP, EP300, BRWD3(2), and PHIP(2)) present an arginine in this position. For an in-depth comparison of the predicted binding mode and the crystal structure, 10 independent runs of explicit solvent molecular dynamics simulations (100 ns each) were carried out starting from the docked ligand conformation which had the inverted orientation of the amide linker. Both orientations of the linker were sampled during the simulations, with that observed in the crystal structure being slightly less populated, possibly due to the fact that the individual trajectories were not long enough to reach equilibrium. The molecular dynamics simulations provided interesting information on the interaction between the carboxylic acid of 10 and the Arg1173 guanidinium. This salt bridge is present only about 70% of the simulation time, which is consistent with the partial solvent exposure of this interaction. In the crystal structure (PDB code 4TQN), the interaction between the carboxylic acid of 10 and the Arg1173 guanidinium is further stabilized by the proximity of the ε-aminon of Lys1130 which belongs to a neighboring bromodomain molecule (shown in magenta in Fig. 3b). Thus, the molecular dynamics simulations supplement the static picture provided by the crystal structure which in this particular case is influenced, at least in part, by the contacts between neighboring protein molecules in the crystalline arrangement.

In conclusion, the ALTA fragment-based docking approach, followed by molecular dynamics simulations, and experimental validation by X-ray crystallography has been successful in the efficient identification of hit compounds whose potency was improved by chemical synthesis of a relatively small set of derivatives. The ligand efficiency and binding selectivity of the acetylbenzene derivatives UZH-1 and UZH-2 [49] compare favorably with CREBBP bromodomain inhibitors reported by others, that is, SGC-CBP30 [51] and I-CBP112 [61] (Fig. 4). It is worth mentioning that, while the inhibitors reported by others have been obtained by

**Figure 4.** Comparison of the optimized ALTA hits for the CREBBP bromodomain and previously reported inhibitors. Heavy atom count (HAC), dissociation constant by means of isothermal titration calorimetry (Kd), and ligand efficiency (LE) are reported for I-CBP112 [61], SGC-CBP30 [51], and the University of Zurich compounds UZH-1 and UZH-2 [49]. The selectivity is calculated as the ratio of the Kd values measured for the CREBBP and BRD4(1) bromodomains via isothermal titration calorimetry (I-CBP112 and SGC-CBP30) or a competition binding assay (UZH-1 and UZH-2).
repurposing BRD4(1) bromodomain ligands, UZH-1 and UZH-2 were generated focusing on the CREBBP bromodomain during all of the steps of our fragment-based high-throughput docking campaigns [48,49].

Conclusions
The ALTA strategy [21,62] starts with the high-throughput docking of a library of fragments followed by the flexible docking of the molecules that contain the top ranking fragments. With this procedure, we have identified small-molecule hits which show low micromolar affinity in vitro and very favorable ligand efficiency for two human bromodomains, BRD4(1) [41] and CREBBP [48]. Importantly, the ALTA procedure is robust with respect to the choices of starting libraries, compound-fragmentation algorithm, docking program, and filtering criteria all of which were significantly different in the two applications to bromodomains [41,48]. In both docking campaigns (as in previous in silico screening campaigns [22]), molecular dynamics simulations started from the pose predicted by docking have played an important role in filtering out compounds with unstable binding mode, and obtaining useful information for hit optimization after in vitro validation.

Most in silico screening campaigns targeting bromodomains have made use of commercial software [20,28-30,33-35,37-40], while the ALTA approach (‘Discovery of BRD4(1) bromodomain inhibitors’ and ‘Discovery of CREBBP bromodomain inhibitors’ sections) employed freely available computer programs for fragment-based high-throughput docking [41,48]. Importantly, the CREBBP ligands identified with the ALTA approach were efficiently optimized to inhibitors with nanomolar potency, favorable ligand efficiency, and selectivity of more than 1000-fold over BRD4(1) (unpublished data). Recently, we have discovered novel hits for two non-BET bromodomains, viz., the BAZ2B [63] and BRPF1 [64] bromodomains, by fragment-based ‘in silico to in crystallo’ strategies. We expect that many more bromodomain hits will be identified in the near future by in silico and/or in vitro fragment-based screening procedures. Furthermore, a subset of them will be optimized into chemical probes targeting individual sub-families of human bromodomains and bromodomains of pathogenic organisms (e.g., Plasmodium falciparum, Leishmania donovani and Trypanosoma brucei, for which bromodomain structures have been solved and released).

Conflict of interest
The authors declare no conflict of interest.

Acknowledgements
We thank Dr. Andrea Unzue, Aymeric Dolbois, and Prof. Cristina Nevado for the synthesis of the derivatives of the acetylbenzene and acetylpiperrole hit compounds. This work was supported in part by a grant of the Swiss Cancer Society (Schweizerische Krebsliga). D.S. is a recipient of the SystemsX.ch translational postdoc fellowship.

References


