Journal of Medicinal Chemistry Article

J. Med. Chem. **2010,** *53,* 7625–7631 **7625**
DOI: 10.1021/jm1007374

Identification of Non-Peptide Malignant Brain Tumor (MBT) Repeat Antagonists by Virtual Screening of Commercially Available Compounds

Dmitri Kireev,* Tim J. Wigle, Jacqueline Norris-Drouin, J. Martin Herold, William P. Janzen, and Stephen V. Frye

Center for Integrative Chemical Biology and Drug Discovery, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Campus Box 7360, 120 Mason Farm Road, Chapel Hill, North Carolina 27599-7363, United States

Received June 18, 2010

The malignant brain tumor (MBT) repeat is an important epigenetic-code "reader" and is functionally associated with differentiation, gene silencing, and tumor suppression.[1−3] Small molecule probes of MBT domains should enable a systematic study of MBT-containing proteins and potentially reveal novel druggable targets. We designed and applied a virtual screening strategy that identified potential MBT antagonists in a large database of commercially available compounds. A small set of virtual hits was purchased and submitted to experimental testing. Nineteen of the purchased compounds showed a specific dose-dependent protein binding and will provide critical structure−activity information for subsequent lead generation and optimization.

Introduction

Epigenetics refers to the study of heritable changes in gene function that occur without a change in the DNA sequence.[4] Epigenetic mechanisms of gene activation and inactivation permit specialization of function between cells even though each cell contains essentially the same genetic code.[5] Typically, changes in the environment trigger post-translational modifications of histone proteins and DNA ("epigenetic marks") including histone lysine and arginine methylation, lysine acetylation, DNA cytosine methylation, histone sumoylation, ubiquitination, ADP-ribosylation, and phosphorylation.[6] These epigenetic modifications form spatial arrangements (often referred to as the "epigenetic code") that recruit protein complexes (epigenetic-code "readers") that cause chromatin to wind and unwind in order to control access of transcription factors to genes. Specific molecular mechanisms by which the "reader" proteins alter gene activation are a subject of intense investigation,[7,8] and small-molecule switches selectively disrupting critical protein−protein interactions would significantly contribute to the ongoing research.

Malignant brain tumor (MBT[a]) domains represent an important class of "code readers" whose function is probably the least understood of this group. From a physiological perspective, MBT proteins are associated with chromatin condensation and act to repress the transcription of genes, ultimately affecting processes such as differentiation, mitotic progression, and tumor suppression.[1−3,9]

Structurally, MBT repeats are similar to the "royal family" histone-binding proteins[10] and recognize prevalently mono- and dimethylated lysines (Kme1 and Kme2).[11] To date, 9 human proteins containing a total of 27 different MBT domains were identified, demonstrating the complex precision with which this specific family of histone binding proteins regulates chromatin accessibility. Therefore, the development of potent and selective small molecule probes for each of the human MBT proteins would facilitate a greater understanding of their roles in stem cell differentiation, cellular reprogramming, and disease etiology. A substantial body of structural information, which is currently available on many MBT domains,[12−16] opens an avenue for rational approaches to the probe-generation effort for this fascinating target class.

Here, we employed a virtual screening strategy to discover non-peptide, cell-penetrant probes for MBT-containing proteins. Indeed, database searching and ligand- or structure-based virtual screening have proved to be useful tools and have become an integral part of the drug discovery process in recent years. The virtual screening process mimics its experimental counterpart and is used to rank or filter large ligand databases in order to yield a compound set "enriched" in hits when experimentally screened. One of the most remarkable virtues of computer-aided approaches is their capacity to screen (i) targets with no assays amenable to an HTS format and (ii) compound collections not readily available for in-house experimental screens. In the search for MBT probes, we screened one the most comprehensive databases of commercially available compounds, the iResearch Library (ChemNavigator),[17] which by the end of 2008 contained more than 50 million procurable chemical samples. To this end, we employed two complementary approaches, one of which consisted of searching for compounds containing Kme1 and Kme2 side chains while the other approach involved sequential application of pharmacophore and docking techniques, hence potentially resulting in more structurally remote compounds mimicking the peptide interaction mode.

A basic prerequisite for an efficient hit discovery process is an accurate, fast, and cost-effective experimental screening

*To whom correspondence should be addressed. Phone: (919) 843-8457. Fax: (919) 843-8465. E-mail: dmitri.kireev@unc.edu.

[a]Abbreviations: MBT, malignant brain tumor; Kme, methylated lysine; HTS, high throughput screening; L3MBTL1, lethal (3) malignant brain tumor-like type 1 protein; MBTD1, malignant brain tumor domain-containing protein 1; H4K20me2, histone 4 dimethylated at lysine 20; OPLS, optimized potentials for liquid simulations; ECFP, extended connectivity fingerprint.

Published on Web 10/08/2010

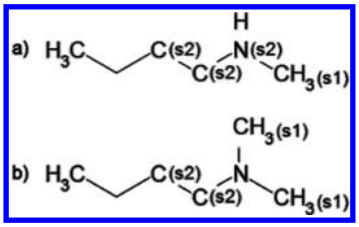

Figure 1. Substructure queries used for searching lysine side chain analogues. The symbols in parentheses specify the number of non-hydrogen substituents to the atom. The (a) and (b) queries respectively code mono- and dimethylated lysine side chains.

technique capable of timely assessment of procured virtual hits. We have previously introduced a novel HTS assay making use of the AlphaScreen technology, and this technique was employed as a primary experimental confirmation for the selected virtual hits.

Materials and Methods

Small-Molecule Data Set. The 2008.2 release of iResearch Library was obtained from ChemNavigator in SD format. Only a 5 967 880 subset of "sourceable" compounds was considered for screening. The structures of these compounds were further cleaned and filtered using the PipelinePilot software.[18] The cleaning protocol included salt stripping, mixture splitting, functional group standardization, and charge neutralization. Ionizable compounds were then converted to their most probable charged forms at pH 7.4 using the LigPrep software.[19] The filtering process included a softened version of the Lipinski rule[20] (2+ violations of Num H-donors, < 6; Num H-acceptors, < 12; MolWeight between 200 and 600; ALogP < 5.5). The filtered set of 5 888 263 compounds ("CHEMNAV_5.9M") was then used for 2D searches and analyses as well as a starting point for the 3D data set generation. PipelinePilot was used for 3D conversion. Stereoisomers were systematically enumerated for chiral compounds with undefined chirality and having less than three chiral centers. For chiral compounds with undefined chirality and having more than two chiral centers a single stereoisomer was produced at random. Compounds with more than 12 rotatable bonds were removed from the 3D set because they represent a substantial burden for both pharmacophore mapping and docking algorithms.

2D Substructure Search. Substructure searches were performed by means of the Pipeline Pilot software on the CHEMNAV_5.9M database. Figure 1 shows the query substructures used in the search for structural analogues of the Kme1 and Kme2 side chains.

Pharmacophore Screening. The pharmacophore was prepared using the Discovery Studio 2.5 software.[18] We made use of the high-resolution (2.05 Å) crystal structure of L3MBTL1 in complex with H4K20me2 (PDB code 2RJF).[13] The Kme2 and adjacent residues were used to define pharmacophoric features encoding three electrostatic-interaction sites (as shown in Figure 2): (i) hydrogen-bond donor (HBD) matching the H4-Lys20 backbone nitrogen interacting with Asn358, (ii) hydrogen-bond acceptor (HBA) of the H4-His18 backbone carboxyl interacting with Asn358, and (iii) amine cation involved in an ionic bond with Asp355. Furthermore, the non-hydrogen atoms of the aromatic residues of L3MBTL1 forming the aromatic cage around the histone−lysine side chain were used to define 16 exclusion spheres. The precision spheres of the pharmacophoric features (i.e., regions of space to which a virtual hit should fit to) were set to 1 Å.

The pharmacophoric screening of the small-molecule set of procurable compounds was performed by means of the Catalyst module of the Discovery Studio 2.5 suite.[18] To this end the 3D SD ligand file was converted to a multiconformer Catalyst database. The conformers were sampled using the BEST method allowing up to 100 conformers per molecule. The enumerated conformers from the Catalyst database were then rigidly fitted against the pharmacophore.

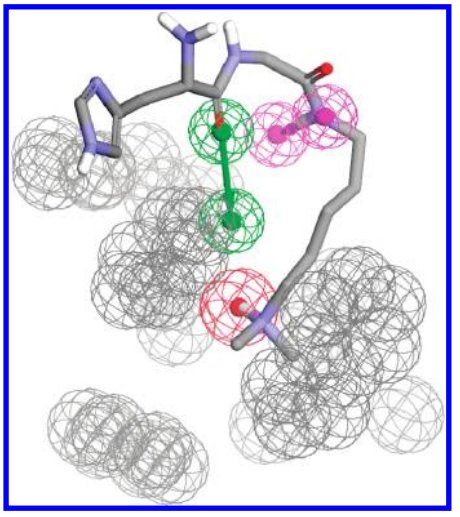

Figure 2. Structure-based pharmacophore model used in the first step of the virtual screening protocol. The structure represents a fragment of the histone's dimethylated Lys20 and His19 (both involved in specific interactions with L3MBTL1) mapped on the pharmacophore model. Two green spheres and an arrow denote a hydrogen-bond acceptor feature (of vector type), two magenta spheres and an arrow designate a hydrogen-bond donor feature, and a red sphere represents an "ionizable positive" feature. The gray spheres are exclusion regions (i.e., regions that cannot be occupied by virtual hits) and derived from L3MBTL1 heavy atoms forming the lysine-binding aromatic cage.

Structure-Based Screening. A high-resolution (2.05 Å) crystal structure of L3MBTL1 in complex with H4K20me2 (PDB code 2RJF)[13] was selected and used at the docking stage. The corresponding PDB file was processed as follows. Hydrogen atoms were added to the protein, the active site was visually inspected, and the appropriate corrections were made for tautomeric states of histidine residues, orientations of hydroxyl groups, and protonation states of basic and acidic residues. The hydrogen atoms were energy-minimized in the MMFF force field[21] using the Macromodel software with the Maestro graphics interface[19] with all the non-hydrogen atoms constrained to their original positions.

Small-molecule structures were docked into the active site of the target protein using the Glide program[19,22] in standard docking precision (Glide SP). The binding region was defined by a 10 Å × 10 Å × 10 Å box centered on the Kme2 side chain of the cocrystallized histone peptide. A scaling factor of 0.8 was applied to the van der Waals radii. Default settings were used for all the remaining parameters. The top 10 poses were generated for each ligand. The docking poses were then energy-minimized with Macromodel in the OPLS2001 force field,[23] with flexible ligand and rigid receptor. The refined poses were reranked on the basis of the calculated interaction energy. The lowest-energy pose for each ligand was selected and rescored in the active site using GlideScore, and the compounds were ranked accordingly.

Hit Analysis and Selection. The Pipeline Pilot software[18] was used for the whole process of hit analysis and selection at all screening steps. Diversity-based selections were generally performed in two steps. First, an automated redundancy reduction is performed by selecting a single representative of a small similarity-based cluster. Compounds in such a cluster should be similar at ≥65% (Tanimoto with ECFP6 fingerprints). In the second step, compounds were clustered into broader families by means of the maximum common substructure (MCS) method. Then 20−50% of compounds were selected from each cluster in such a way that larger clusters contributed smaller percentages.

The output ligands were aligned to their respective MCS to facilitate an ad hoc selection.

Quality Control of Compounds Screened. The quality control of the plate containing the screened compounds was performed by diluting a 1 μL DMSO stock solution (100 μM) with 29 μL of MeOH. The sealed plate was directly used to inject 5 μL for each well. HPLC data of all compounds were acquired using an Agilent 6110 series system with the UV detector set to 220 nm. Samples were injected onto an Agilent Eclipse Plus 4.6 mm × 50 mm, 1.8 μM, C18 column at room temperature. A mobile phase of A being H_2O + 0.1% acetic acid and B being MeOH + 0.1% acetic acid was used. A linear gradient from 10% to 100% B in 5.0 min was followed by pumping 100% B for another 2 min with a flow rate of 1.0 mL/min. Mass spectral data were acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source.

The purity of the compounds screened is found to be 95% or higher.

AlphaScreen Assay. The constructs containing MBT repeats for L3MBTL1 (residues 200−530), L3MBTL3 (residues 225−555), L3MBTL4 (residues 44−371), and MBTD1 (residues 130−566) cloned into a pET28a-MHL plasmid and transfected into BL21 DE3 *E. coli* were generously provided by the Structural Genomics Consortium and purified as previously described.[24] The following additional peptides were synthesized and purified by high-performance liquid chromatography (HPLC) by the Tufts Peptide Synthesis Core Facility (Boston, MA) to act as substrates for L3MBTL3, L3MBTL4, and MBTD1. A peptide representative of monomethyl lysine 36 on histone H2A (H2AK36me) with the sequence biotin-AHA-GRVHRLLRK(me)GNYSER-COOH was used as a substrate for L3MBTL3 and L3MBTL4, and a peptide representative of H4K20Me with the sequence biotin-AHA-KGGAKRHRK(me)VLRDNIQ-COOH was used as a substrate for MBTD1. Here and further in the text, "(me)" denotes the site of the monomethylated lysine, AHA indicates the inclusion of a 6-aminohexyl linker between the N-terminal residue and the biotin group, and COOH indicates a free carboxylic acid on the C-terminus.

Compounds for the dose-response runs were resuspended to 100 mM in DMSO in barcoded glass vials and sonicated using a Covaris XX (Covaris, Woburn, MA). The compounds were plated as 3-fold dilutions over 10 points using a Tecan Genesis (Tecan, Männedorf, Switzerland) in 384-well V-bottom polypropylene microplates (Greiner, Monroe, NC). A Multimek NS X-1536 fitted with a 384-channel head (Nanoscreen, Charleston, SC) was used to spot 1 μL of the compounds into 384-well polypropylene V-bottom microplates that were sealed and stored at −20 °C. On the day of use, the compounds were prepared for screening by diluting 100-fold in 1× assay buffer and 1 μL of the diluted titrations were spotted into 384-well Proxiplates to which 9 μL of protein and peptide cocktail was added to initiate the assay. The AlphaScreen assay was performed as previously described for L3MBTL1[24] with the following modifications for screening the other MBT proteins. L3MBTL3 was assayed at 200 nM with 150 nM H2AK36me1. L3MBTL4 was assayed at 100 nM with 150 nM H2AK36Me, and MBTD1 was assayed at 100 nM with 150 nM H4Kme20. The binding of L3MBTL1, L3MBTL3, and MBTD1 to their cognate peptides was detected using 5 μg/mL AlphaScreen nickel chelate acceptor and strepatvidin donor beads, and the interaction between L3MBTL4 and its cognate peptide was detected using 10 μg/mL of the same beads. Dose-response runs were analyzed using ScreenAble software (Screening Solutions LLC, Chapel Hill, NC), and IC_{50} values were calculated using four-parameter fits or using three-parameter fixed top fits as necessary. The counter-screen was performed to identify any compound interference of AlphaScreen signal transduction as previously described[24] after the compounds were prepared as described above.

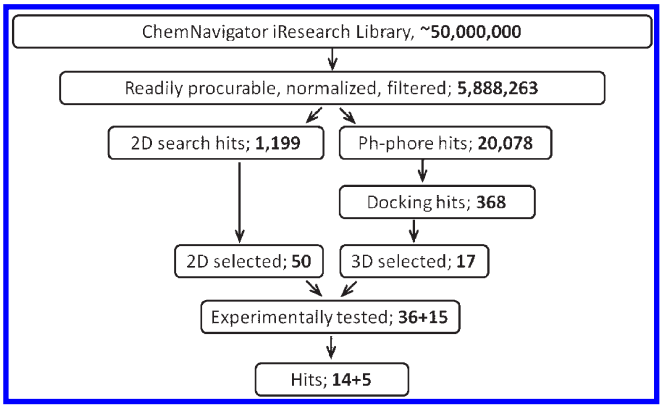

Figure 3. Overall screening process chart. Numbers in boxes are counts of compounds in the specified category. "2D selected" and "3D selected" are respective outcomes of automated and ad hoc selections as described in the Materials and Methods. Numbers in the boxes "Experimentally tested" and "Hits" designate compounds coming from, respectively, 2D substructure- and 3D structure-based screening.

Results and Discussion

The overall screening process is outlined in Figure 3. We decided to process 2D substructure searches and 3D structure-based virtual screening as two parallel threads. The rationale for this choice was to combine hits from an ad hoc approach based on a medicinal chemist's judgment with those from a computational approach taking direct advantage of the available protein structure. More specifically, the ad hoc approach may provide ligands whose binding mode and affinity cannot be adequately assessed by virtual screening techniques. Alternatively, a structure-based approach yields hits along with a sound hypothesis about their binding mode, thus allowing immediate guidance to structural modifications that may improve potency.

MBT domains represent a unique class of methyllysine binders. For instance, unlike most other domains belonging to the royal family and recognizing Kme3, MBT binds lower methylation states (i.e., Kme1 and Kme2). Moreover, MBTs recognize their respective histone methylation sites employing a "cavity-insertion" mode, which buries the Kme side chain within a deep cleft, as opposed to a sequence-dependent "surface-groove" mode, involving a wider methyllysine-binding pocket.[8] MBT domains have a highly conserved architecture, an "aromatic cage, including Phe379, Trp382, and Tyr386 (numbering for L3MBTL1), as shown in Figure 4.

These aromatic residues are involved in cation−π interactions with the methylated ammonium group, while a highly conserved Asp355 forms an ionic bond and is critical for the lower methyl mark recognition. For instance, in three human MBT domains known to bind Kme (D2-hL3MBTL1, D2-hSCML2, and D4-L3MBTL2),[25] Phe379, Trp382, and Asp355 are conserved in all of them while Tyr386 is conserved in two domains (mutated to Phe in D2-hSCML2). Hence, given the high degree of the pocket conservation, we have chosen hL3MBTL1 as a representative member of the MBT family for the current virtual screening study, expecting that some of identified virtual hits will also be active on other family members.

The critical importance of the Kme cavity insertion combined with the relatively low impact of peptide sequence[26] prompted us to start our hit fishing with a minimalist hypothesis that a set of close methyllysine side chain mimics might

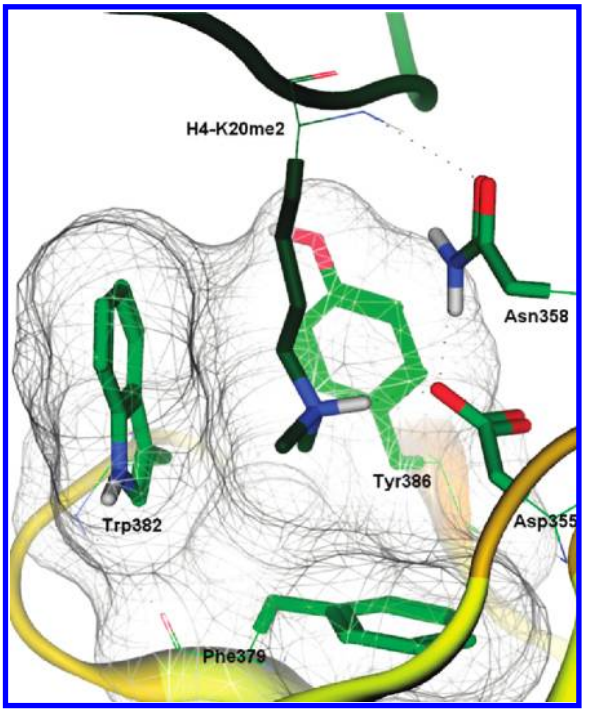

Figure 4. Methyllysine binding site of hL3MBTL1 (light green) in complex with an H4 histone 10-mer peptide (dark green) [PDB code 2RJF]. The surface (light gray) outlines the methyllysine-insertion cavity.

be a good starting point for an experimental screening study with some hope that the non-Kme mimic portion of the molecules selected would serendipitously provide additional binding interactions. Searching CHEMNAV_5.9M using C_3CH_2NMe and $C_3CH_2NMe_2$ (see Figure 1) as substructure queries resulted in 1199 hits. We then applied a redundancy reduction procedure that consisted of clustering of the hits obtained into very compact (in terms of internal similarity) clusters and selecting one central compound per cluster. The resulting 344 cluster centers were grouped into 288 families featuring common Murcko frameworks.[27] The families were then subjected to an ad hoc selection, based on consideration of a combination of physical and structural properties that determine their leadlike potential. This analysis yielded 35 compounds, and some of them were supplemented by close structural analogues that resulted in a final list of 50 compounds. Some of those compounds were further excluded from the list based on price and, upon purchase, on QC analysis, which resulted in an experimentally tested set of 36 methyllysine analogues.

In addition to the substructure search with restrictive queries, we also intended to take a more direct advantage of the crystal structure. However, we estimated that application of a docking method to 5 million compounds would not be an adequate solution. Indeed, in the absence of a diverse set of known binders, it would not be possible to validate the docking/scoring protocol, leading to a higher rate of false positives, particularly because the relatively shallow binding cavity will only be partially occupied by the majority of ligands, which would still be highly scored because of their propensity to readily form geometrically accurate hydrogen bonds with solvent-exposed residues.[28,29]

Alternatively, a pharmacophore approach enables the identification of ligands possessing functional features characteristic of an active compound, implying that they bind the target similarly to the prototypic active. Therefore, a pharmacophore may serve as an efficient filter to select ligands that are likely to bind in a similar fashion to the histone peptide in X-ray structures. Docking/scoring pharmacophore hits in the protein binding site will then play a complementary role for an accurate assessment of steric and van der Waals interactions.

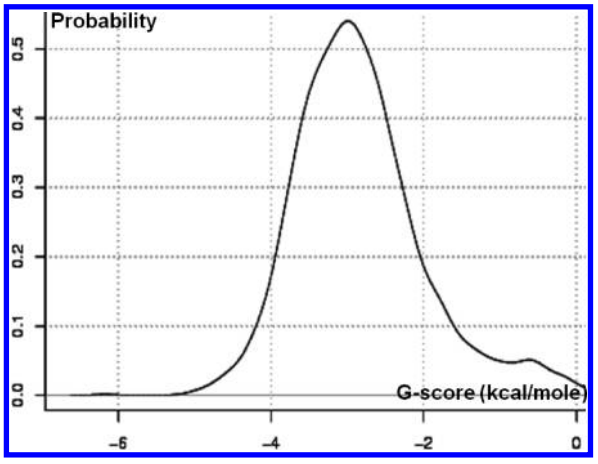

Figure 5. Probability density distribution extrapolated from G-score values resulting from docking of 10 000 random compounds having physical properties similar to those of pharmacophore hits. Although this set may accidentally contain some actives, the distribution as a whole represents that of inactive compounds.

The pharmacophore model was built using the crystal structure of L3MBTL1 in complex with a cocrystallized histone peptide [2RJF] (as described in Materials and Methods). Pharmacophore screening of CHEMNAV_5.9 M resulted in 20 078 hits, which represents an affordable workload for the downstream docking/scoring without any additional filtering. Docking of pharmacophore hits was performed using Glide at Standard Precision as described in Materials and Methods. A total of 60 126 poses (20 055 ligands) had a G-score of < 0 kJ/mol. To be consistent with the rationale of a sequential pharmacophore-docking protocol, we retained only those 16 830 poses (8947 ligands) that interact with Asp355 and Asn358 (interactions that our pharmacophore model is based upon). In order to choose a statistically significant G-score cutoff, we made use of the probability density distribution of G-score values obtained by docking a set of 10 000 decoys. These decoys were randomly selected from 334 992 commercially available compounds, having physical profiles similar to those of pharmacophore hits (i.e., one positive ionizable group, ≥ 2 HBA, and ≥ 1 HBD). Our assumption was that a random selection from a broad compound set would have a distribution of G-scores characteristic of that of inactive compounds and would be indicative of the false positive rate at a given G-score value. On the basis of the clearly asymmetric nature of these distributions, we did not assume any analytical form and made use of a nonparametric, kernel density estimator (with Gaussian kernels). The distribution (see Figure 5) shows that inactive compounds are quite unlikely to have a G-score of > 5.5 kcal/mol when interacting with the binding site of L3MBTL1, and therefore, this value may be set as a threshold to select docking hits.

The 168 primary hits (with G-score of > 5.5 kcal/mol) were clustered into families of structurally related compounds. Poses of the best scored representatives of each of 36 clusters were reviewed within the protein binding site. Only poses having at least 2 hydrogen bonds, in addition to the required

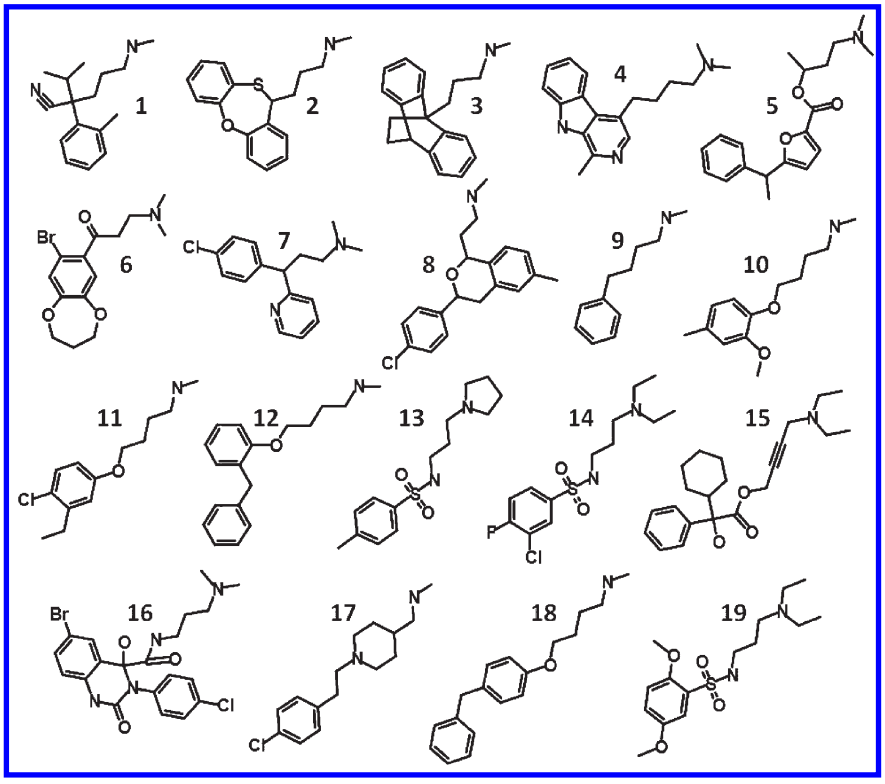

Figure 6. Structures of 19 virtual hits confirmed by AlphaScreen.

Table 1. IC_{50} Results for the Alpha-screen Hits against Four MBT-Containing Proteins[a]

N	origin	counterscreen [% at 100 μM]	L3MBTL1 [μM]	L3MBTL3 [μM]	L3MBTL4 [μM]	MBTD1 [μM]
1	2D	3	18 ± 2.4 (3)	> 100	> 100	> 100
2	2D	9	14 ± 0.69 (2)	ND	> 100	> 100
3	2D	20	61 ± 20 (16)	85 ± 5.3 (3)	> 100	31
4	2D	15	21 ± 2.1 (3)	5.7 ± 1.6 (3)	> 100	73 ± 14 (2)
5	2D	22	84 ± 1.3 (2)	75 ± 1.4 (2)	> 100	66 ± 7.1 (3)
6	2D	35	> 100	> 100	12 ± 2.0 (2)	> 100
7	2D	1	41 ± 10 (3)	42 ± 0.39 (3)	> 100	34 ± 9.3 (2)
8	2D	23	> 100	ND	78 ± 6.5 (2)	> 100
9	2D	5	68 ± 5.0 (3)	> 100	> 100	> 100
10	2D	30	> 100	84 ± 13 (2)	92 ± 7.2 (2)	> 100
11	2D	−1	96 ± 1.1 (2)	ND	> 100	> 100
12	2D	8	91 ± 1.4 (2)	ND	> 100	> 100
13	SB	−5	17 ± 0.13 (2)	57 ± 11 (3)	> 100	59 ± 5.1 (3)
14	SB	8	> 100	30 ± 5.7 (3)	> 100	> 100
15	SB	29	> 100	54 ± 5.2 (3)	> 100	> 100
16	SB	3	52 ± 10 (2)	ND	> 100	37 ± 4.2 (3)
17	2D	−3	69 ± 2.4 (2)	ND	> 100	> 100
18	2D	8	91 ± 1.4 (2)	> 100	> 100	> 100
19	SB	−6	> 100	48 ± 20 (3)	> 100	> 100

[a] The "origin" column indicates whether a hit was identified by substsructural searches (2D) or by the structure-based approach (SB). In the case of multiple dose-response experiments, mean ± standard deviation (number of experiments) values are given. ND means that no dose dependent response was observed for these compounds.

ionic bond, with the protein were retained. Finally, 17 ligands representative of 4 clusters were selected as candidates for purchase.

In the end, a total of 51 compounds (36 resulting from 2D search and 15 from virtual screening) were actually screened against our panel of 4 MBT-containing proteins, i.e., L3MBTL1, L3MBTL3, L3MBTL4, MBTD1 (see Supporting Information for a complete SD file of experimentally tested samples). To make sure that even weakly active compounds are identified, all 51 compounds were submitted to dose-response AlphaScreen experiments in a concentration range of 5 nM to 100 μM. Nineteen of 51 tested compounds (Figure 6) demonstrated an unambiguous dose-dependent effect in this assay (Table 1). Fourteen of these experimental hits are part of the 36 compounds identified by the substructure search, while 5 come from the set of 15 pharmacophore/docking hits. Figure 7 shows the dose-response curves and structures of the most potent compounds from each hit category as well as the highest scored pose of the most potent docking hit having a pyrrolidine moiety interacting with Asp355 (in place of mono- or dimethylated ammonium). The two hit categories are complementary in terms of their potential for future

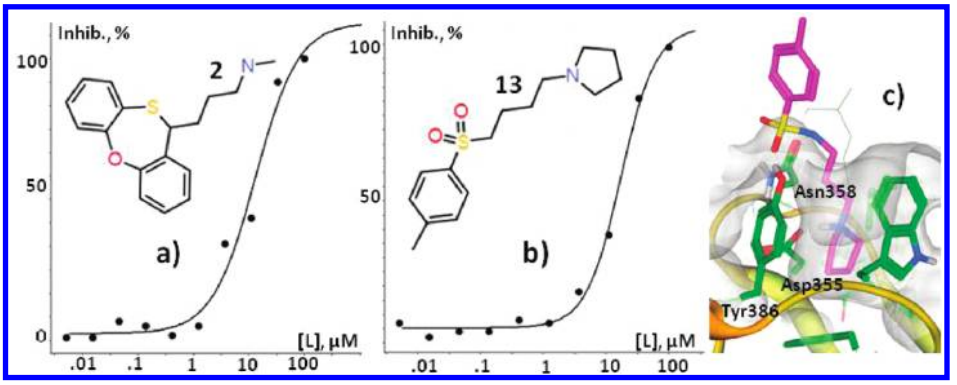

Figure 7. (a, b) Alpha-screen dose-response curves for two most potent hits coming from substructure searches (compound **2**) and the pharmacophore/docking protocol (compound **13**). (c) The highest scored docking pose (magenta sticks) of compound **13** is superposed with the crystal structure of the methylated lysine residue (dark green lines). The sulfonamide group of **13** forms hydrogen bonds with the side chains of Tyr386 and Asn358, and the protonated nitrogen of the pyrolidine group forms an ionic bond with Asp355.

chemical optimization. For example, some of the most potent 2D hits (**1**, **2**, and **6**) selectively bind to a single MBT-containing protein from our panel. However, the binding mode of these hits cannot be reliably hypothesized and many of them cannot be mapped to our pharmacophore model. Conversely, the pharmacophore/docking hits may be readily mapped to the pharmacophore, and thus, their binding mode to most of MBT domains may be hypothesized with high confidence. It still remains unexplained why the structure-based hits are selective to one or two of four MBT-containing proteins on our screening panel despite that they all possess a pharmacophore, which should confer an ability to bind any MBT domain. This selectivity is reassuring in light of future chemical optimization, and its structural rationale will certainly be understood when more ample structure−activity data are available. The structure-based hits also provide evidence that Kme1 or Kme2 moieties are not the only functional groups capable of binding the MBT aromatic cage. For example, the pyrolidine-containing compound **13** is one of the most potent (IC_{50} = 17 μM) among the screened compounds. Also, compound **15**, which shows some activity against L3MBTL3 (IC_{50} = 54 μM), has a rigid alkyne linker instead of a lysine-like alkane chain.

It is noteworthy that one of the substructure-search hits is maprotiline (**3**), an approved drug and strong norepinephrine uptake inhibitor, also active against a broad set of aminergic G-protein-coupled receptors (GPCR). Consequently, maprotiline, in addition to its known biological properties, may also have some chromatin-related activity, although the affinity to L3MBTL1 is 3 orders of magnitude lower than the affinity to its primary target and may be of little pharmacological relevance.

The overall SAR for identified hits from both categories is quite flat (5.7−96 μM) and may be explained by the current binding mode hypothesis which implies that a large portion of each hit molecule is exposed to solvent. Additionally, the potency of currently identified hits is certainly insufficient to consider them as probe[30] candidates and will be the subject to further chemical optimization. The upcoming optimization will target a more substantial "burying" of a ligand in the MBT binding pocket. Possible directions would include modifications of the linker between the deeply buried amino group and the outer aromatic motif as well as ortho substitutions on the outer aromatic group (e.g., ortho substituted compound **13** analogues).

Conclusions

In silico approaches have matured to become an established source of novel and diverse chemical tools to study and exploit the pharmaceutical potential of novel biological targets. Here we applied a combination of computational techniques in order to identify small-molecule ligands for MBT-containing proteins. MBT domains constitute a novel class of chromatin regulators, epigenetic-code "readers", associated with chromatin condensation and gene repression, ultimately affecting processes such as differentiation, mitotic progression, and tumor suppression.[1−3,9]

In this report, we have made use of two parallel and complementary strategies: (i) ad hoc substructure searches for ligands possessing a lysine-like fragment, potentially resulting in structurally diverse hits with unexpected binding modes, (ii) a semiautomated sequential protocol involving 3D pharmacophores and structure-based screening to detect hits whose binding mode mimics that of endogenous ligands thus providing structural insights to subsequent potency improvements.

Both strategies produced plausible hit hypotheses leading to the purchase and experimental testing of the most promising compounds. We applied a recently developed screening technique,[24] making use of the AlphaScreen technology, to assess the potency of virtual hits against a panel of 4 MBT-containing proteins. A total of 19 tractable MBT antagonists, coming from both 2D and structure-based screening protocols, showed specific dose-dependent effects in the AlphaScreen assay.

After appropriate optimization, these hits may provide a basis to study the biological function as well as pharmaceutical potential of MBT-containing proteins as a new target class.

Acknowledgment. We thank Structural Genomics Consortium, Toronto, Canada, for providing protein constructs. We also thank Dr. Duane Bronson for assistance with ScreenAble software. This research was supported by startup funds to SVF provided by the Carolina Partnership and by Grant RC1-GM090732-01 from the National Institutes of Health.

Supporting Information Available: Zipped files containing chemical structures and biological activities of experimentally tested compounds, in Excel and SD formats. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Trojer, P.; Li, G.; Sims, R. J., 3rd; Vaquero, A.; Kalakonda, N.; Boccuni, P.; Lee, D.; Erdjument-Bromage, H.; Tempst, P.; Nimer, S. D.; Wang, Y. H.; Reinberg, D. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **2007**, *129* (5), 915–928.

(2) Grimm, C.; de Ayala Alonso, A. G.; Rybin, V.; Steuerwald, U.; Ly-Hartig, N.; Fischle, W.; Muller, J.; Muller, C. W. Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. *EMBO Rep.* **2007**, *8* (11), 1031–1037.

(3) Wu, S.; Trievel, R. C.; Rice, J. C. Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. *FEBS Lett.* **2007**, *581* (17), 3289–3296.

(4) Berger, S. L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23* (7), 781–783.

(5) Bernstein, B. E.; Meissner, A.; Lander, E. S. The mammalian epigenome. *Cell* **2007**, *128* (4), 669–681.

(6) Gelato, K. A.; Fischle, W. Role of histone modifications in defining chromatin structure and function. *Biol. Chem.* **2008**, *389* (4), 353–363.

(7) Campos, E. I.; Reinberg, D. Histones: annotating chromatin. *Annu. Rev. Genet.* **2009**, *43*, 559–599.

(8) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025–1040.

(9) Bornemann, D.; Miller, E.; Simon, J. Expression and properties of wild-type and mutant forms of the Drosophila sex comb on midleg (SCM) repressor protein. *Genetics* **1998**, *150* (2), 675–686.

(10) Maurer-Stroh, S.; Dickens, N. J.; Hughes-Davies, L.; Kouzarides, T.; Eisenhaber, F.; Ponting, C. P. The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **2003**, *28* (2), 69–74.

(11) Adams-Cioaba, M. A.; Min, J. Structure and function of histone methylation binding proteins. *Biochem. Cell Biol.* **2009**, *87* (1), 93–105.

(12) Eryilmaz, J.; Pan, P.; Amaya, M. F.; Allali-Hassani, A.; Dong, A.; Adams-Cioaba, M. A.; MacKenzie, F.; Vedadi, M.; Min, J. Structural studies of a four-MBT repeat protein MBTD1. *PLoS One* **2009**, *4* (10), e7274–e7274.

(13) Min, J.; Allali-Hassani, A.; Nady, N.; Qi, C.; Ouyang, H.; Liu, Y.; MacKenzie, F.; Vedadi, M.; Arrowsmith, C. H. L3MBTL1 recognition of mono- and dimethylated histones. *Nat. Struct. Mol. Biol.* **2007**, *14* (12), 1229–1230.

(14) Li, H.; Fischle, W.; Wang, W.; Duncan, E. M.; Liang, L.; Murakami-Ishibe, S.; Allis, C. D.; Patel, D. J. Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* **2007**, *28* (4), 677–691.

(15) Guo, Y.; Nady, N.; Qi, C.; Allali-Hassani, A.; Zhu, H.; Pan, P.; Adams-Cioaba, M. A.; Amaya, M. F.; Dong, A.; Vedadi, M.; Schapira, M.; Read, R. J.; Arrowsmith, C. H.; Min, J. Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic Acids Res.* **2009**, *37* (7), 2204–2210.

(16) Santiveri, C. M.; Lechtenberg, B. C.; Allen, M. D.; Sathyamurthy, A.; Jaulent, A. M.; Freund, S. M.; Bycroft, M. The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J. Mol. Biol.* **2008**, *382* (5), 1107–1112.

(17) ChemNavigator. www.chemnavigator.com.

(18) Accelrys Software Inc. www.accelrys.com.

(19) Schrödinger, LLC. www.schrodinger.com.

(20) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44* (1), 235–249.

(21) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490–519.

(22) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.

(23) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Ttesting of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.

(24) Wigle, T. J.; Herold, J. M.; Senisterra, G. A.; Vedadi, M.; Kireev, D. B.; Arrowsmith, C. H.; Frye, S. V.; Janzen, W. P. Screening for inhibitors of low-affinity epigenetic peptide−protein interactions: an AlphaScreen-based assay for antagonists of methyl-lysine binding proteins. *J. Biomol. Screening* **2010**, *15* (1), 62–71.

(25) Bonasio, R.; Lecona, E.; Reinberg, D. MBT domain proteins in development and disease. *Semin. Cell Dev. Biol.* **2009**, *21* (2), 221–230.

(26) Trojer, P.; Reinberg, D. Beyond histone methyl-lysine binding: how malignant brain tumor (MBT) protein L3MBTL1 impacts chromatin structure. *Cell Cycle* **2008**, *7* (5), 578–585.

(27) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893.

(28) Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model.* **2003**, *9* (1), 47–57.

(29) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44* (7), 1035–1042.

(30) Frye, S. V. The art of the chemical probe. *Nat. Chem. Biol.* **2010**, *6* (3), 159–161.