



Published in final edited form as:

J Am Chem Soc. 2011 April 13; 133(14): 5357–5362. doi:10.1021/ja110432e.

Biophysical probes reveal a “compromise” nature of the methyllysine binding pocket in L3MBTL1

Cen Gao, J. Martin Herold, Dmitri Kireev, Tim Wigle, Jacqueline L. Norris, and Stephen Frye

Center for Integrative Chemical Biology and Drug Discovery, Division of Medicinal Chemistry and Natural Products, UNC Eshelman School of Pharmacy, University of North Carolina

Abstract

Histone lysine methylation (Kme) encodes essential information modulating many biological processes including gene expression and transcriptional regulation. However, the atomic-level recognition mechanisms of methylated histones by their respective adaptor proteins are still elusive. For instance, it is unclear how L3MBTL1, a methyl-lysine histone code reader, recognizes equally well both mono- and di-methyl marks, but ignores unmodified and trimethylated lysine residues. We made use of Molecular Dynamics (MD) and Free Energy Perturbation (FEP) techniques in order to investigate the energetics and dynamics of the methyllysine recognition. Isothermal Titration Calorimetry (ITC) was employed to experimentally validate the computational findings. Both computational and experimental methods were applied to a set of designed “biophysical” probes that mimic the shape of a single lysine residue and reproduce the binding affinities of cognate histone peptides. Our results suggest that, besides forming favorable interactions, the L3MBTL1 binding pocket energetically penalizes both methylation states and has most probably evolved as a “compromise” that non-optimally fit to both mono- and di-methyl-lysine marks.

Introduction

Histone lysine-methylation plays a key role in transcriptional regulation¹. Abnormal methylation patterns may lead to various pathologies including cancer^{2,3}. At the molecular level, methylation marks act via recruitment of protein complexes that mediate transcriptional activation or repression⁴. MBT repeats form the smallest and the least studied class of “chromatin readers”, i.e. protein modules that bind to methyl-lysine marks on histone tails. Functionally, these proteins localize to chromatin and regulate transcription by a currently unknown molecular mechanism⁵. For instance, L3MBTL1, a protein that features three MBT domains, recognizes mono- and di-methyl-lysine marks on H1.4K26, H3K4, H3K9, H3K27 and H4K20 *in vitro* and associates with Heterochromatin Protein 1 (HP1 γ), and the Retinoblastoma protein (Rb)⁶. These interactions result in a transcriptionally nonpermissive chromatin structure *in vitro* and the negative regulation of multiple genes through the E2F/Rb oncogenic pathway^{7,8}. Remarkably, despite the complexity of the L3MBTL1 interaction network, it has been demonstrated that transcriptional repression by L3MBTL1 relies upon a single methylation mark (H4K20me1)⁹. Moreover, the lower

Corresponding author: Dmitri Kireev Center for Integrative Chemical Biology and Drug Discovery University of North Carolina at Chapel Hill Campus Box 7363, Genetic Medicine Building, room 2097 120 Mason Farm Road Chapel Hill, NC, 27599-7363 Office: (919) 843-8457 Fax: (919) 843-8465 dmitri.kireev@unc.edu .

Supporting Information: Supplementary Schemes; Supplementary Tables; Binding Curves for ITC Experiments; Experimental Procedures for all new compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

methylation state specificity of L3MBTL1 was exclusively mediated by the second MBT domain⁹.

Although the biological implications of the various histone methylation states are subject to intense investigations⁵, the atomic-scale energetics and dynamics of Kme recognition by respective histone code readers remain elusive. Yet, exogenous modulation of the histone code using small molecule agents will require a detailed understanding of how lysine methylation states are recognized by their respective adaptor proteins. More specifically, it is unclear how adding a single methyl group to Kme0 or removing it from Kme3 results in a huge gain in affinity to a “reader” protein, while adding a methyl to Kme1 does not affect the affinity at all.

Free Energy Perturbation (FEP)¹⁰, coupled to Molecular Dynamics (MD) simulations, is well suited to study energetics and dynamics at the atomic level. Moreover, valuable additional information, such as electrostatic or van der Waals (vdW) contributions to the free energy of binding can also be extracted from computational simulations. Obtaining this valuable information requires a judicious choice of a study system. Ideally, the potential energetic changes should be confined to the interaction between the methylated lysine side chain and the “reading” pocket of L3MBTL1. Although straightforward *in silico*, this approach may appear less biologically relevant than previously reported experimental studies^{6,7,11} involving 10-15 residue histone fragments. However, a substantial body of evidence demonstrates that such a reductionist hypothesis (that biological function relies upon a localized pocket-residue interaction) is applicable in the context of Kme recognition by L3MBTL1. For instance, it has been shown that the MBT-Kme recognition occurs in a histone sequence independent manner^{6,7,11}. Moreover, available X-ray structures of MBT-containing proteins demonstrate a “cavity-insertion” recognition mode, with the modified lysine side-chain fully buried within the MBT binding cage⁴, while the rest of the histone chain is exposed to the solvent and forms very low-impact interactions with the reader protein. Finally, the reductionist hypothesis is also supported by cell-based data. More specifically, it has demonstrated that repressive function of L3MBTL1 relies solely on its second MBT repeat and does not need its interaction partners (i.e. Rb and HP1 γ)⁹. From the computational perspective, the use of a mono-peptide-like probe whose interactions would be confined to the Kme binding pocket is also highly preferred because allows to avoid any parasitic interactions that might occur between L3MBTL1 and a polypeptide probe. Moreover, the smaller probes will have significant phase space overlaps, which will result in a better convergence of FEP calculations¹².

Here, in line with the above rationale, we investigate the atom-level energetics of MBT-Kme recognition in a system where L3MBTL1 interacts with small molecules that mimic a single lysine residue. Six compounds that we refer to as biophysical probes were synthesized. The compounds **1-4** (Table 1; Supporting Information) demonstrate the same affinity profile towards L3MBTL1 as cognate peptides do. To gain a deeper understanding of the L3MBTL1 binding pocket architecture we have also synthesized and studied two exogenous alkylated lysine analogs, compounds **5** (N(me)et) and **6** (pyrrolidine). In order to corroborate computational findings with experimentally observable data, we measured the affinity of the designed biophysical probes to L3MBTL1 by means of Isothermal Titration Calorimetry (ITC).

Results and Discussion

At the first step, we studied binding of all six synthesized compounds to a fragment of L3MBTL1 consisting of three MBT domains (residues 200-522) by means of ITC. The results (Table 1, Supplementary Figures) suggest that the probes, with exception of

compounds **1** (Nme0) and **4** (Nme3), demonstrate a dose-dependent interaction with affinities in the order pyrrolidine > Nme2 \approx N(me)et \approx Nme1 \gg Nme3 \approx Nme0. The relative order and absolute binding free energies of the studied compounds are consistent with the trend observed for histone peptides, which supports their use as biophysical probes to study MBT-Kme recognition. In order to make sure that our probes do actually bind to the second MBT domain of L3MBTL1 we have also measured binding of compounds **2**, **3**, **5** and **6** to the D355A L3MBTL1 mutant. This mutation has previously been demonstrated to efficiently switch off any binding of mono- or di-methylated histones⁶ because of the absence of the main binding anchor Asp355. Our ITC experiments with the mutant did not show any measurable binding (Supplementary ITC figures). This strongly supports the hypothesis that the studied compounds bind to the lysine binding pocket of the second MBT domain.

Our experimental affinity data confirm the intriguing structure-activity relationships (SAR), where the same chemical modification, i.e. adding or removal of a single methyl group, in a similar context, may result either in no change in affinity or in its full loss. More specifically, it is unclear how adding a single methyl group to Nme0 or removing it from Nme3 results in a huge gain in affinity to a “reader” protein, but does not affect the affinity when adding a methyl to Nme1. Consequently, we performed a series of FEP and MD simulations (see Methods and Supplementary Scheme 1a) in order to provide an atomic-scale structural rationale for these “atypical” SAR. The computed free energies obtained demonstrate a strong correlation with ITC results, which justifies the use of FEP for the energetic analysis of Kme recognition.

The same computational protocol was also used to determine relative weight of polar and non-polar contributions to the binding affinity and to ascertain “preferred” interaction modes for each compound. To this end, we made use of virtual probes that represent neutral isomorphs of compounds **1-6** obtained by replacing their amino nitrogens with carbon atoms. As previously demonstrated¹³⁻¹⁵, the binding free energy difference between an ionizable compound and its respective non-polar isomorph ($\Delta\Delta G_{c-i}^{FEP}$) can be exclusively attributed to polar interactions, i.e. hydrogen bonding, cation- π and long range ionic interactions (see Methods and Supplementary Scheme. 1b). Furthermore, the nonpolar contribution to the difference in affinities of two compounds can be expressed as the difference in affinities of the two respective nonpolar isomorphs. Here, we performed six additional FEP calculations to compute relative binding affinities of neutral isomorphs **1'-6'** to L3MBTL1 (see Table 1) and used these affinities in the further energetic and structural analyses.

We first compared data obtained for compounds **2** (Nme1) and **3** (Nme2). Their respective affinities to L3MBTL1 were similar in both ITC and FEP experiments. However, binding free energies of their respective neutral isomorphs **2'** and **3'** are significantly different (Scheme 1, Supplementary Scheme 2). The neutral isomorph **2'** binds tighter than **3'** by 1.10 kcal/mol due to its more efficient non-polar interactions. Additionally, we made use of conventional MD simulations to provide a structural interpretation for the MBT-Nme recognition. Ten thousand structural snapshots from 20 ns simulations were clustered and analyzed in order to determine the most representative bound ligand conformations. As shown in Figure 1a, the bound state of compound **2** (Nme1) may be represented by 5 conformations, varying significantly in positions of both the amino nitrogen and the methyl group. However, despite its high mobility within the binding pocket, the only methyl group of compound **2** manages to keep optimal distances of *ca.* 3.5 Å to each of three aromatic side chains, Phe379, Trp382 and Tyr386 (see Methods and Fig. 2). In contrast, the bound compound **3** (Nme2) is represented by three quite distinct orientations of its methyl groups (Fig. 1b). The distance distribution in Fig. 2 shows that only one methyl group can keep an

optimal distance to the aromatic side chains, while the other can significantly diverge beyond the range of optimal vdW interactions. One of representative conformers of compound **3** is identical to the X-ray conformation of Kme2 with L3MBTL1¹¹ and two others differ only in methyl group orientations while keeping the amino nitrogen in a position of optimal electrostatic interaction with Asp355.

Therefore, both structurally and energetically, the identical binding affinities of **2** (Nme1) and **3** (Nme2) are achieved through quite different recognition mechanisms (as depicted in Fig. 3). Most remarkably, both methylation forms are moderate-affinity MBT binders. For instance, compound **2** (Nme1) benefits from more efficient vdW interactions and a higher bound-state mobility, leading to a favorable entropic contribution. However, it is penalized by suboptimal electrostatic interactions because its positive charge is delocalized between two hydrogens of which only one can efficiently interact with the carboxyl group of Asp355. Alternatively, the compound **3** (Nme2) is penalized by less favorable non-polar interactions, but benefits from a much stronger electrostatic contribution due to its highly localized positive charge. Consequently, it appears that the lysine pocket of L3MBTL1 that equally non-optimally binds both cognate ligands (i.e. Kme1 and Kme2) has emerged as an evolutionary “compromise” between two possible pocket designs, each of which would tightly bind either Kme1 or Kme2. Energetically, this “compromise” pocket equally penalizes both methylation states, although through different interaction forces.

The proposed “compromise” pocket hypothesis is also compatible with the extremely weak affinity of compounds **1** (Nme0) and **4** (Nme3). Indeed, compounds **1** (Nme0) lacks all favorable vdW interactions, from which its closest analog **2** (Nme1) benefits. This results in a non-polar free energy loss of 2.90 kcal/mol. Furthermore, compound **1** (Nme0) is also penalized by a weaker electrostatic contribution because its positive charge is delocalized between three hydrogen atoms, resulting in a loss of an additional 1.83 kcal/mol. As to the low affinity of compound **4** (Nme3), it has been previously hypothesized that steric repulsion is the major force preventing MBT domains from binding the trimethylated lysine⁴. Our decomposition analysis suggests that while the non-polar term (0.73 kcal/mol) penalizes compound **4**, the polar term (1.38 kcal/mol) is even more important for the decline of its affinity (relative to **3**). The above is in agreement with the observation that a negatively charged residue is essential for a functional MBT domain, but not mandatory for a Kme3 binding domain⁴. For example, the Kme3 binding pocket of BPTF, a PHD finger-containing protein, does not contain any acidic residues; interestingly, a Y17E mutation in the binding pocket led to BPTF preferentially recognizing Kme2 over Kme3⁶.

Finally, in line with the finding that the native methylation state binding is imperfect by design, one might also infer that a tighter MBT binder is possible. To test this hypothesis we have synthesized and studied two alkylated lysine analogs **5** (N(me)et) and **6** (pyrrolidine), whose ability to bind MBT domains was established in our previous studies^{16,17}. ITC data show that compound **6** is the most potent of all alkylation states we have examined and binds tighter than **3** (Nme2) by 0.66 kcal/mol, while the potency of compound **5** (N(me)et) is comparable to that of **3** (Nme2). Our free energy decomposition analysis suggests that the calculated gain in affinity of **6** (pyrrolidine) over **3** (Nme2) (1.51 kcal/mol), is due to a huge electrostatic contribution of 2.75 kcal/mol. Compound **6** is however penalized for its two extra carbons by a non-polar penalty of 1.23 kcal/mol compared to **3**. On the atomic level, the significant electrostatics-related advantage of **6** is hard to explain considering a relatively modest change in partial charge as compared to **3** (Nme2). To further examine this interaction, we analyzed the conventional MD trajectory to provide a structural rationale for the apparent electrostatics reward. The analysis demonstrates that the pyrrolidine ring, slightly withdrawn from the pocket, has a very low mobility in its bound state and is constrained to optimally interact with Asp355 (Fig. 1d). Moreover, the pyrrolidine ring is

parallel to the aromatic side chain of Trp382 and forms an aromatic-aliphatic ring stacking motif, commonly seen in the Cambridge Structure Database¹⁸ and in some proteins featuring an aromatic binding cage¹⁹.

To conclude, in this study we proposed the L3MBTL1 lysine binding pocket has evolved as a “compromise” that features equally non-optimal fits to both mono- and demethylation states. This finding may help in understanding the evolutionary development of the methylation machinery and also provide guidance to the development of more efficient exogenous modulators of the histone code.

Experimental section

Synthesis

Detailed synthetic procedures and characterization are described in the supporting information.

Protein and ligand structure preparation for MD and FEP

We used the second MBT domain (residues 366 to 424) of the X-ray structure of L3MBTL1 at 2.05 Å resolution (PDB entry: 2RJF)¹¹ as a starting point for all MD/FEP simulations. The protein structure was prepared using Protein Preparation Wizard from the Maestro software suite²⁰. Hydrogen atoms were added to the structure and the whole system was minimized using the OPLS force field²¹. Three-dimensional structures of small-molecule ligands were generated with Ligprep²⁰, and multiple low energy conformations were saved. To obtain initial binding poses, all ligands were docked to the protein using Glide’s Standard Precision (SP) mode²². These binding poses were visually inspected and the most reasonable structure was selected as the initial bound structure for dynamic simulation. No acceptable docking pose was found for compound **4** (Nme3). Its initial bound structure was obtained by superimposing it to compound **3** (Nme2), followed by a local minimization. All simulation systems were solvated in an orthorhombic TIP3P water box with a minimum of 10 Å between the box boundary and any solute atom.

Molecular Dynamics

Molecular dynamics was employed to sample the system ensembles during FEP calculation. AMBER99SB force field²³ was used to describe the receptor. This force field has been proven to efficiently reproduce the cation- π interaction²⁴, which is crucial for our model system. It is also recently applied to investigate the methyl-lysine dependent binding of H3 histone tails to the HP1 chromodomain¹⁴. General Amber Force Field (GAFF)²⁵ was used to model ligand molecules. Atomic partial charges were derived by restrained electrostatic potential (RESP)²⁶ method, using geometries optimized at HF/6-31G* level in Gaussian 03²⁷.

All MD simulations were conducted using Desmond v2.2^{28,29}. Particle mesh Ewald (PME) summation was used to treat long-term interactions³⁰. SHAKE algorithm³¹ was used to restrain all bonds involving hydrogen atoms. The system was energy-minimized and then gradually heated to 300 K during 0.5 ns. An additional 1.0 ns simulation at 300 K was performed to further equilibrate the system. All production runs were performed with the NPT ensemble. System temperature and pressure were regulated by Langevin thermostat and barostat³². The integration time steps were 1 fs, 1 fs, 3 fs for, respectively, the bonded term, the near term (short-range electrostatic and van der Waals interactions) and the far term (long range electrostatic interactions).

Free Energy Perturbation

Binding free energy differences ($\Delta\Delta G_{i \rightarrow j}$) for two similar ligands i and j that bind to the same protein were calculated by means of the FEP method according to the thermodynamic cycle shown in Supplementary Scheme 1a. Relative binding affinities $\Delta\Delta G_{i \rightarrow j}$ was computed by calculating the free energy difference of two ligands in solution and in protein, where:

$$\Delta\Delta G_{i \rightarrow j} = \Delta G_j - \Delta G_i = \Delta G_{i \rightarrow j}^{\text{complex}} - \Delta G_{i \rightarrow j}^{\text{solution}}$$

Alchemical mutations were performed on a dual topology basis, comprising a series of intermediate states. To speed up the convergence, soft-core potential was used to scale down the vdW potential. For a simulation window k , the hybrid Hamiltonian was the sum of Lennard-Jones, electrostatic and bonded interaction terms:

$$H(x, \lambda_k) = H_{LJ}(x, \lambda_{k,vdWA}, \lambda_{k,vdWB}) + H_{elec}(x, \lambda_{k,elecA}, \lambda_{k,elecB}) + H_{bonded}(x, \lambda_{k,bondA}, \lambda_{k,bondB})$$

where λ_k is a coupling parameter between the initial and final Hamiltonians. Most FEP simulations involved a total of 17 windows, with an 8.0 ns production run in each window. The λ_k values used in our calculations were listed in Supplementary Table 1. To ensure a smooth alchemical transformation from compounds **5** (N(me)et) to **6** (pyrrolidine), simulations were performed on a 23 window basis, with three additional windows in the beginning of the simulation and three in the end.

For each system, two independent calculations with different random seeds were performed. Results from both simulations were combined and subjected to the Bennett acceptance ratio (BAR) analysis³³ in order to compute free energy differences between every two adjacent simulation windows. The block bootstrapping method was used to estimate the statistical uncertainties³⁴. In this study, 20 block bootstrapping trials were conducted and the standard errors over these trials were reported in Supplementary Table 2.

Free energy decomposition

The binding site of the second MBT domain of L3MBTL1 features an aromatic cage and an aspartate residue. Hence, the interaction with cationic lysine-like ligands may involve ionic, hydrogen bonding, cation- π and vdW interactions. MD/FEP simulations allow to determine the energetic components of the ligand-protein binding. More specifically, virtual non-polar isomorphs of the actual ligands have proved a useful tool to decompose the relative free energy of binding $\Delta\Delta G_{i \rightarrow j}$ into electrostatic and non-polar contributions^{14,35}. The electrostatic term mainly includes hydrogen bonding, cation- π and long range ionic interactions, while the non-polar contribution is often referred to as a charge-independent term, including van der Waals interactions and the hydrophobic effect. Here we performed six FEP calculations to compute the relative binding affinities between our polar compounds **1-6** and their respective non-polar isomorphs **1'-6'** (e.g., the non-polar isomorph for Nme3 is Cme3). These calculations completed the thermodynamic cycle showed in Supplementary Scheme 1b. Because the nonpolar compounds have an identical size and shape to their corresponding polar counterparts, the relative binding free energy between each two neutral isomorphs ($\Delta\Delta G_{i' \rightarrow j'}$ in Supplementary Scheme 1b) can be used to estimate the non-polar effect during the binding of the charged compounds i and j . On the other hand, the difference between $\Delta\Delta G_{i' \rightarrow j'}$ and the relative binding free energy of the polar compounds $\Delta\Delta G_{i \rightarrow j}$ can be considered as the electrostatic contribution ($\Delta\Delta G_{i \rightarrow j} - \Delta\Delta G_{i' \rightarrow j'}$ in

Supplementary Scheme 1b). The decomposed free energies are summarized in Scheme 1 and Supplementary Scheme 2.

Representative MD conformations

In order to provide structural insights into the Kme recognition mechanism we selected representative ligand-protein conformations from MD simulations. To this end, we performed conventional 20 ns MD simulations for each protein-ligand complex. System coordinates were saved every 2 ps, yielding 10,000 snapshots per MD run. The snapshots were then clustered based on the distances from the ligand's heavy atoms to the aromatic residues that constitute the active site. Clustering was performed using Pipeline Pilot v7.5³⁶. The final representative structures were obtained by a visual inspection of all cluster centers.

General Procedure for Isothermal Titration Calorimetry (ITC) Experiments

For the ITC measurements, L3MBTL1 was extensively dialyzed into ITC buffer (20 mM Tris-HCl, pH 8, 25 mM NaCl and 2 mM β -mercaptoethanol). Subsequently, the concentration was spectrometrically established; the extinction coefficient ϵ for L3MBTL1 is 90870. The ITC experiments were performed at 25°C, using an AutoITC₂₀₀ microcalorimeter (MicroCal Inc., USA). Experiments were performed by injecting 1.5 μ L of a 1 mM solution of the compounds into a 200 μ L sample cell containing 50 μ M or 100 μ M L3MBTL1: all weakly binding compounds **2**, **3** and **5** were titrated to a 100 μ M protein solution, compounds **1**, **4** and **6** was titrated to 50 μ M protein. A total of 26 injections were performed with a spacing of 180 seconds and a reference power of 8 μ cal/s. Compounds were dissolved in ITC buffer at 10 mM and were diluted to 1 mM. A control experiment for each compound was also performed and the heat of dilution was measured by titrating each compound into the buffer alone. The heat of dilution generated by the compounds was subtracted, and the binding isotherms were plotted and analyzed using Origin Software (MicroCal Inc., USA). The ITC measurements were fit to a one-site binding model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

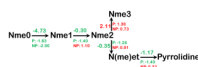
Acknowledgments

We thank Structural Genomics Consortium, Toronto for providing protein constructs. This research was supported by startup funds for SVF provided by the Carolina Partnership and by the grants RC1-GM090732-01 from the National Institutes of Health and the Innovation Award from the University Cancer Research Fund. We also thank Dr. Harry Stern (University of Rochester) for helpful discussions.

References

- (1). Zhang Y, Reinberg D. *Genes Dev.* 2001; 15:2343–2360. [PubMed: 11562345]
- (2). Rivenbark AG, Coleman WB, Stahl BD. *FASEB J.* 2009; 23:38.1.
- (3). Kondo Y, Shen L, Issa J-P. *J. Mol. Cell. Biol.* 2003; 23:206–215.
- (4). Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. *Nature Struct. Mol. Biol.* 2007; 14:1025–1040. [PubMed: 17984965]
- (5). Bonasio R, Lecona E, Reinberg D. *Semin Cell Dev Biol.* 2009; 21:221–230. [PubMed: 19778625]
- (6). Li H, Fischle W, Wang W, Duncan EM, Liang L, Murakami-Ishibe S, Allis CD, Patel DJ. *Mol. Cell.* 2007; 28:677–691. [PubMed: 18042461]
- (7). Trojer P, Li G, Sims RJ 3rd, Vaquero A, Kalakonda N, Bocconi P, Lee D, Erdjument-Bromage H, Tempst P, Nimer SD, Wang YH, Reinberg D. *Cell.* 2007; 129:915–928. [PubMed: 17540172]
- (8). Nevins JR. *Hum. Mol. Genet.* 2001; 10:699–703. [PubMed: 11257102]

- (9). Kalakonda N, Fischle W, Boccuni P, Gurvich N, Hoya-Arias R, Zhao X, Miyata Y, Macgrogan D, Zhang J, Sims JK, Rice JC, Nimer SD. *Oncogene*. 2008; 27:4293–4304. [PubMed: 18408754]
- (10). Chipot, C.; Pohorille, A. *Free energy calculations : theory and applications in chemistry and biology*. Study ed. Springer; New York: 2007.
- (11). Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, MacKenzie F, Vedadi M, Arrowsmith CH. *Nat Struct Mol Biol*. 2007; 14:1229–1230. [PubMed: 18026117]
- (12). Pohorille A, Jarzynski C, Chipot C. *J Phys Chem B*. 2011; 114:10235–10253. [PubMed: 20701361]
- (13). Ma JC, Dougherty DA. *Chem Rev*. 1997; 97:1303–1324. [PubMed: 11851453]
- (14). Lu Z, Lai J, Zhang Y. *J. Am. Chem. Soc*. 2009; 131:14928–14931. [PubMed: 19788305]
- (15). Hughes RM, Wiggins KR, Khorasanizadeh S, Waters ML. *Proc Natl Acad Sci U S A*. 2007; 104:11184–11188. [PubMed: 17581885]
- (16). Kireev D, Wigle TJ, Norris-Drouin J, Herold JM, Janzen WP, Frye SV. *J Med Chem*. 2010; 53:7625–7631. [PubMed: 20931980]
- (17). Herold JM, Wigle TJ, Norris JL, Lam R, Korboukh VK, Gao C, Ingerman LA, Kireev DB, Senisterra G, Vedadi M, Tripathy A, Brown PJ, Arrowsmith CH, Jin J, Janzen WP, Frye SV. *J Med Chem*. 2011 (in press).
- (18). Paul P. *Crystal engineering*. 2002; 5:3–8.
- (19). Hopkins AL, Ren J, Milton J, Hazen RJ, Chan JH, Stuart DI, Stammers DK. *J Med Chem*. 2004; 47:5912–5922. [PubMed: 15537346]
- (20). Maestro Suite. Schrodinger LLC; 2009.
- (21). Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc*. 1996; 118:11225–11236.
- (22). Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. *J Med Chem*. 2004; 47:1750–9. [PubMed: 15027866]
- (23). Kini RM, Evans HJ. *J Biomol Struct Dyn*. 1991; 9:475–88. [PubMed: 1687724]
- (24). Felder C, Jiang H-L, Zhu W-L, Chen K-X, Silman I, Botti SA, Sussman JL. *J Phys Chem A*. 2001; 105:1326–1333.
- (25). Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. *J Comput Chem*. 2004; 25:1157–74. [PubMed: 15116359]
- (26). Cornell WD, Cieplak P, Bayly CI, Kollmann PA. *J Am Chem Soc*. 1993; 115:9620–9631.
- (27). Frisch, MJ.; Trucks, GW.; Schlegel, HB. *Gaussian 03*. 2004. Revision C.02
- (28). Desmond Molecular Dynamics System. version 2.2. D. E. Shaw Research; 2009.
- (29). Kevin, JB.; Edmond, C.; Huafeng, X.; Ron, OD.; Michael, PE.; Brent, AG.; John, LK.; Istvan, K.; Mark, AM.; Federico, DS.; John, KS.; Yibing, S.; David, ES., editors. 2006. p. 43-43.
- (30). Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. *J Chem Phys*. 1995; 103:8577–8593.
- (31). Andersen H. *J Comp Phys*. 1983; 52:24–34.
- (32). Pastor RW, Brooks BR, Szabo A. *Mol Phys*. 1988; 65:1409–1419.
- (33). Bennett CH. *J Comp Phys*. 1976; 22:245–268.
- (34). Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK. *J Mol Biol*. 2009; 394:747–763. [PubMed: 19782087]
- (35). Gallivan JP, Dougherty DA. *Proc Natl Acad Sci U S A*. 1999; 96:9459–9464. [PubMed: 10449714]
- (36). Pipeline Pilot. ver. 7.5. Accelrys Software Inc.; 2009.

**Scheme 1.**

Free energy decomposition for compounds **1-6** (as annotated in Table 1). P / NP refer respectively to polar and non-polar contribution to $\Delta\Delta G^{\text{FEP}}$.

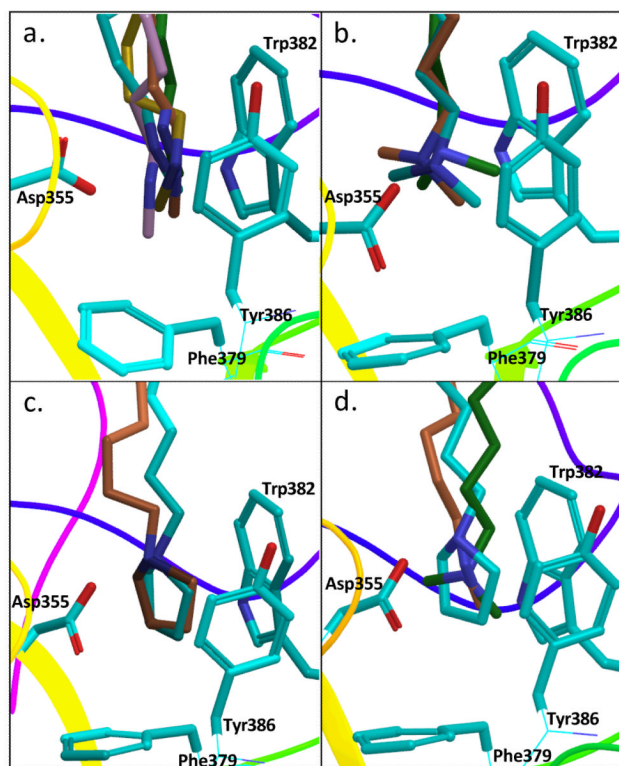


Figure 1. Representative conformers of compounds **2**, **3** and **6** derived from a 20 ns MD simulation in the Kme binding site of L3MBTL1: a) compound **2** (Nme1) demonstrates a high conformational diversity, b) compound **3** (Kme2) features 3 distinct orientations of its methyl groups, c) the pyrrolidine group of **6** is fixed in the cage allowing only for the carbon chain flexibility, d) superposition of most frequent orientations of Nme1, Nme2 and pyrrolidine heads.

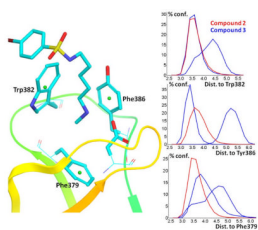


Figure 2. Distance distributions between the methyl groups of biophysical probes and aromatic side chains of the binding pocket of L3MBTL1. The only methyl group of compound **2** (Nme1, red curve) is most of the time in the distance range corresponding to the optimal vdW interactions. Only one of two methyl groups of compound **3** (Nme2, blue curves) manages to keep an energetically favorable distance to the aromatic side chains.

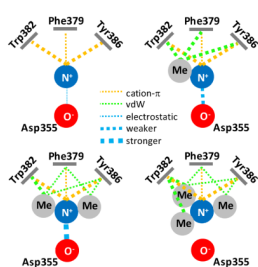
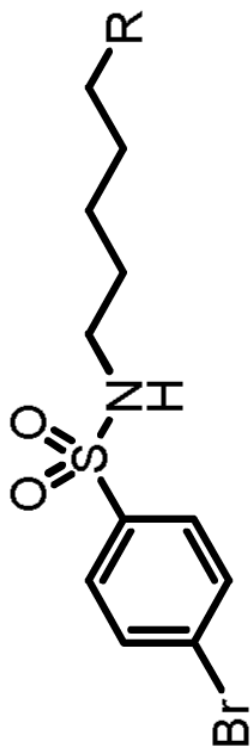


Figure 3.

A qualitative recognition model of various lysine methylation states by MBT domains. The Nme0 state demonstrates weak interactions to all binding site residues, while Nme3 is lacking a critical interaction with Asp355. The Nme1 and Nme2 states display the same binding affinity, but quite different interaction patterns.

Table 1

Six designed biophysical probes, their binding affinities to L3MBTL1 and respective computed energetic contributions.



N	ID	R	K_d (μ M)	$\Delta\Delta G_{c \rightarrow 2}^{Exp}$	$\Delta\Delta G_{c \rightarrow 2}^{FEP}$	Isomorph	R'	$\Delta\Delta G_{c \rightarrow 1}^{FEP}$
1	UNC587	Nme0	>10000	>2.7	4.73	1'	Cme0	-1.28
2	UNC588	Nme1	104 \pm 10	0.00	0.00	2'	Cme1	0.55
3	UNC589	Nme2	73 \pm 19	-0.21	-0.30	3'	Cme2	1.95
4	UNC590	Nme3	>10000	>2.7	1.81	4'	Cme3	0.57
5	UNC592	N(me)et	91 \pm 11	-0.08	-0.65	5'	C(me)et	3.21
6	UNC591	Pyrrolidine	21 \pm 2	-0.94	-1.82	6'	Cyclopentane	4.70

Dissociation constants K_d are measured by ITC and are the average of two independent measurements. Free energy values are expressed in kcal/mol. Binding free energies calculated by FEP ($\Delta\Delta G_{c \rightarrow 2}^{FEP}$) are given relative to the experimental binding free energy of compound 2 (-5.38 kcal/mol). The R' column denotes the neutral isomorphs of compounds 1-6. $\Delta\Delta G_{c \rightarrow 1}^{FEP}$ is the binding free energy difference between a compound and its neutral isomorph. Compounds 1, 4 did not demonstrate any detectable binding in our ITC experiments (Supplementary ITC data).