Assessment of Free Energy Predictors for Ligand Binding to a Methyllysine Histone Code Reader

Cen Gao, **J. Martin Herold**, and **Dmitri Kireev**[*]
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

Methyllysine histone code readers constitute a new promising group of potential drug targets. For instance, L3MBTL1, a Malignant Brain Tumor (MBT) protein, selectively binds mono- and di-methyllysine epigenetic marks (KMe, KMe_2) that eventually results in the negative regulation of multiple genes through the E2F/Rb oncogenic pathway. There is a pressing need in potent and selective small-molecule probes that would enable further target validation and might become therapeutic leads. Such an endeavor would require efficient tools to assess the free energy of protein-ligand binding. However, due to an unparalleled function of the MBT binding pocket (i.e. selective binding to KMe/KMe_2) and because of its distinctive structure representing a small aromatic "cage", an accurate assessment of its binding affinity to a ligand appears to be a challenging task. Here, we report a comparative analysis of computationally affordable affinity predictors applied to a set of seven small-molecule ligands interacting with L3MBTL1. The analysis deals with novel ligands and targets, but applies widespread computational approaches and intuitive comparison metrics that makes this study compatible with and incremental to earlier large scale accounts on the efficiency of affinity predictors. Ultimately, this study has revealed three top performers, far ahead of the other techniques, including two scoring functions, PMF04 and PLP, along with a simulation-based method Molecular Mechanics Poisson-Boltzmann/Surface Area (MM-PB/SA). We discuss why some methods may perform better than others on this target class, the limits of their application, as well as how the efficiency of the most CPU-demanding techniques could be optimized.

Introduction

Epigenetic mechanisms are responsible for the cell differentiation and phenotypic changes under the influence of the environment[1]. These mechanisms are triggered by a set of post-translational modifications ("epigenetic marks") of histone proteins and DNA[2]. Among those modifications, histone methylation plays a key role in transcriptional regulation. Abnormal methylation patterns may lead to various pathologies including cancer[1]. At the molecular level, the spatial arrangements of methylation marks (Kme), i.e. the Kme histone code, act via recruitment of protein complexes (Kme code "readers") that mediate transcriptional activation or repression[3]. Malignant Brain Tumor (MBT) repeats that bind to methyllysine marks on histone tails[4] form the smallest and the least studied class of "chromatin readers". Functionally, these proteins localize to chromatin and regulate transcription by a currently unknown molecular mechanism. For instance, L3MBTL1, a protein that features three MBT domains, recognizes mono- and di-methyllysine marks on H1.4K26, H3K4, H3K9, H3K27 and H4K20 in vitro and associates with Heterochromatin

Corresponding author: Dmitri Kireev, Center for Integrative Chemical Biology and Drug Discovery, The 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.

Protein 1 (HP1γ), and the Retinoblastoma protein (Rb)[5]. It has been demonstrated that the Kme recognition by L3MBTL1 occurs in a histone sequence independent manner and is exclusively mediated by the second MBT domain[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], which makes L3MBTL1 and related MBT-containing proteins a group of attractive therapeutic targets. Hence, there is a pressing need in potent and selective small-molecule probes that would enable further target validation and might become leads for therapeutic optimization. Several recent publications have reported small-molecule MBT antagonists[8–10]. However, their binding affinities are relatively weak and need further improvement.

Computational approaches can provide guidance for the optimization of the current hits and help to identify novel chemical series[11]. Moreover, recent successes in structure determination for multiple Kme readers open the avenue for the structure-based design and virtual screening[12]. Such an endeavor will require efficient tools for the assessment of the free energy of protein-ligand binding. However, an accurate assessment of its binding affinity to a novel ligand may appear to be a challenging task. This challenge is dual: first, the binding pocket is quite small (i.e. just enough to accommodate a lysine side chain),, which means that a substantial portion of ligand might be solvent-exposed[13]; second, the pocket demonstrates a very intriguing ligand selectivity pattern, i.e. recognizes equally well both mono- and di-methyl marks (Kme1/Kme2), but completely ignores unmodified (Kme0) and trimethylated (Kme3) lysine residues. Recently, we have studied how L3MBTL1 discriminates Kme1/Kme2 from Kme0 and Kme3 by means of Molecular Dynamics and Free Energy Perturbation (FEP) which allowed to predict the relative binding free energies for the studied ligands[9]. This study demonstrated that the binding free energies can in principle be accurately calculated for this class of proteins. However, when it comes to virtual screening and structure-based design, FEP would not be a viable free-energy prediction option. First, its computational cost is very high, especially when scaled to hundreds of ligand-protein complexes. Second, alchemical transformations between ligands as required by FEP are difficult to automate and can only be applied to relatively similar ligands. Therefore, there is a need for computationally efficient predictors of the ligand-protein binding free energy for the histone Kme readers, which constitute a novel class of potential drug targets.

Here, we report a comparative analysis of computationally affordable free energy predictors applied to a set of seven small-molecule ligands interacting with L3MBTL1. A relatively small number of ligands in the test set is because quite few MBT binders have been published so far. However, these compounds cover a broad range of binding affinities (from >−2.32 to −7.23 kcal/mol) that is an absolute prerequisite for such a benchmark analysis. The evaluated predictors were chosen in order to represent various alternative approaches to free energy calculation. We considered the following the method categories. The first category includes dynamics-based Molecular Mechanics-Poisson Boltzmann/Surface Area (MM-PB/SA), Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) and Linear Interaction Energy (LIE). These methods are computationally intensive and allow a throughput of about 10 ligand-protein complexes a day. Another category includes single point versions of MM-PB/SA, MM-GB/SA and LIE which do not involve any conformational sampling and make use of a single conformation, most often obtained from a fast docking algorithm. These techniques allow to process hundreds of ligands a day and are generally used as a post-screening rescoring solution. Finally, we have also decided to give a try to a few scoring functions that are often embedded in docking engines and used for ranking docking poses and virtual hits. Here again, we selected a representative subset of popular algorithms. For instance, Dock6[14] and Autodock4[15] are illustrative of the force-

field-based approach, PMF[16], PMF04[17] are knowledge-based, while PLP[18], LigScore[19], Jain[20] and GlideSP[21] make use of various empirical-based schemas.

We believe that this study is complementary to earlier large scale analyses involving multiple protein targets and small-molecule ligands such as those performed by Wang *et al*[22] and Warren *et al*[23]. To this end, we have made use of widespread free energy predictors with standard parameter settings as well as of intuitive comparison metrics such as Pearson's R^2, Kendall's τ_b and Predictive Index (PI).

Materials and Methods

System Setup and Molecular Dynamics Simulation

Molecular system was set up in a similar way as described previously in our paper[9] and is only briefly summarized here. The starting protein structure is the second MBT domain (residues 366–424) of L3MBTL1 (PDB entry: 2RJF). The protein was treated with Protein Preparation Wizard module of Maestro[21]. Initial bound structures were obtained by superimposing them to the X-ray structure of the dimethylated lysine residue in complex with L3MBTL1 (PDB entry: 2RJF), followed by a local minimization. The initial bound structure of compound **7** was obtained from its own X-ray structure (PDB entry: 3P8H)[10]. 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 dynamic simulations were performed using Desmond molecular dynamics program V2.2 [24,25]. Particle mesh Ewald (PME) summation[26,27] was used to treat long-term interactions. SHAKE algorithm[28,29] was used to restrain all bonds involving hydrogen atoms. The system was minimized and then gradually heated to 300 K in 0.5 ns. An additional 0.3 ns simulation at 300 K was performed to further equilibrate the system. Two independent 20 ns NPT production runs were performed for each receptor-ligand complexes. System temperature and pressure were regulated by Langevin thermostat and barostat [30,31]. The integration time step was 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). Structure snapshots were saved every 10 ps. Finally, as required for LIE method, each free-state ligand was also simulated in solution with a similar procedure.

MM-PB/SA and MM-GB/SA

In MM-PB/SA and MM-GB/SA schemes, the free energy of binding is estimated as:

$$\Delta G_{bind} \approx \langle \Delta E_{MM} \rangle + \langle \Delta G_{solv} \rangle + \langle \Delta G_{SA} \rangle - \langle T\Delta S_{conf} \rangle \tag{1}$$

Here the molecular mechanics term ΔE_{MM} is the difference between the energy of the complex and the energies of free-state ligand and receptor in gas phase. ΔG_{solv} and ΔG_{SA} are the electrostatic solvation energy and surface area energy change upon ligand binding. In MM-PB/SA approach, the polar solvation term is computed by iteratively solving Poisson-Boltzmann (PB) equation, while in MM-GB/SA this term is calculated with the linearized Generalized Born (GB) model. The last term $T\Delta S_{conf}$ is the correction accounting for the conformational entropy loss of the ligand, computed from either normal node or quasiharmonic analysis. This term is often neglected for structurally similar ligands. The symbol $\langle \dots \rangle$ denotes the ensemble average over a series of conformations, normally generated by either Molecular Dynamics (MD) or Monte Carlo (MC) simulations. In this study, a "single-trajectory" approach[32] was used. MM-PB/SA and MM-GB/SA calculations were performed using mm_pbsa.pl as implemented in AMBER 10[33]. MM-PB/SA rescoring

was conducted using ZAP[34]. MM-GB/SA rescoring was computed using Prime MM-GB/SA module[21].

Linear Interaction Energy (LIE)

LIE method estimates the binding free energy based on a linear electrostatic response approximation and an empirical term accounting for non-polar interactions [35].

$$\Delta G_{bind} \approx \alpha \left(\langle E_{l-s}^{vdW} \rangle_{bound} - \langle E_{l-s}^{vdW} \rangle_{free} \right) + \beta \left(\langle E_{l-s}^{elec} \rangle_{bound} - \langle E_{l-s}^{elec} \rangle_{free} \right) + \gamma \tag{2}$$

Here 'l-s' refers to the interaction energy between the ligand and its surroundings. The two energy terms were computed from MD trajectories in explicit solvent. α and β are two adjustable scaling factors, while γ is a constant close to zero. The actual values of α and β vary in different publications and depend on the modeling system[36,37]. The values recommended by the authors of the approach are $\alpha \approx 0.18$ and β in the range from 1/3 to 1/2[37]. In the current study, a "vrun" simulation was performed in Desmond in order to extract the interaction energy between ligand and its surroundings. Similarly to MM-GB/SA, LIE was later extended to a simplified rescoring process by incorporating a continuum solvent model[38] and by utilizing a single energy minimization[39]. Here we used the Surface Generalized Born (SGB)-LIE rescoring model as implemented in Liaison[21]. In this model, the electrostatic contribution for the net free energy is computed by the sum of Coulomb interaction energy (E_{coul}) and the SGB-solvent reaction field energy (E_{rxn}).

Statistical criteria

The predictive power of evaluated methods was assessed by three complementary statistical criteria: 1) Pearson's R^2, 2) Kendall's τ_b and 3) the Predictive Index (PI). Kendall's τ_b is a rank-order correlation coefficient and is defined as:

$$\tau_b = \frac{n_c - n_d}{\sqrt{\left(\frac{n(n-1)}{2} - t \right) \left(\frac{n(n-1)}{2} - u \right)}} \tag{3}$$

where n_c is the number of concordant pairs of ranks; n_d is the number of discordant pairs of ranks. t and u are the number of tied values in experimental ΔG and predicted ΔG, respectively.

Predictive index (PI) is a different indicator of rank ordering first described by Pearlman and Charifson[40]. It is defined as:

$$PI = \left(\sum_{j>i} |E_i - E_j| c_{ij} \right) / \left(\sum_{j>i} |E_i - E_j| \right) \tag{4}$$

where

$$c_{ij} = \begin{cases} 1, & (E_i - E_j)/(P_i - P_j) > 0 \\ -1, & (E_i - E_j)/(P_i - P_j) < 0 \\ 0, & P_i - P_j = 0 \end{cases} \tag{5}$$

Here E_i is the experimental value of ligand i and P_i is the predicted value. The value of PI ranges from −1 to 1, where 1 means a perfect rank prediction and −1 means a completely wrong prediction. It is a more sophisticated ranking measure than the Kendall's τ_b because it

also includes a weighting term that depends on the differences between the consecutive experimental values.

Results and Discussion

Compound set and experimental data

We selected a total of seven compounds (Table 1) in order to test affordable predictors of binding free energy to the aromatic cage of L3MBTL1, a representative of methyllysine reader proteins. Compounds **1** to **4** were designed to mimic four endogenous methylation states, i.e. N, NMe, NMe_2 and NMe_3. Compounds **5** (N(Me)Et) and **6** (Pyrrolidine) are two exogenous alkylated lysine analogs, which bind equally well or better than compounds **2** and **3**. Further optimization of compound **6** led to compound **7**, which contains a restrained linker[10]. Remarkably, despite quite subtle structural differences, these compounds showed a broad range of binding affinity from >20000 to 5 μM (see Table 1). Five compounds, with exception of compounds **1** (N) and **4** (NMe_3), exhibited a dose-dependent interaction with L3MBTL1 as measured by Isothermal Titration Calorimetry (ITC). The binding affinities were in the order of **7** > **6** (Pyrrolidine) > **3** (NMe_2) ≈ **5** (NMe(Et)) ≈ **2** (NMe) ≫ **4** (NMe_3) ≈ **1** (N).

Free energy predictions

Computational results of our comparative analysis are summarized in Table 2. The results are arranged in three blocks. In the first block, we considered the most computationally intensive, dynamics-based techniques, i.e. PB/SA, GB/SA and LIE. Here, FEP results obtained from our last study are also listed in this block for comparison purposes. The second block includes single-point versions of the first-block methods that do not require dynamics simulations. Finally, the third block comprises 10 popular scoring functions employed by popular virtual screening software. Three popular metrics have been chosen to express the performance of each method: (i) coefficient of determination or Pearson's R^2 is a fraction of variance of experimental data that is explained by predicted data; (ii) Kendall's τ_b reflects the similarity of the orderings ranked respectively by experimental and predicted data; and (iii) predictive index (PI) is a hybrid metrics that quantify both similarities in rank orderings and the gaps between consecutive experimental values. Each of the three metrics can be used to rank the predictors from best to worst. As it can be seen from Table 2, all three ranks converge significantly for three affinity prediction methods – PMF04, PLP and MM-PB/SA – that can therefore be regarded as top performers in this study. More generally, we can reasonably assume that all three assessment metrics are important and hence if any of three coefficients rank a method as a poor performer, this method should tend toward the bottom of the overall ranking. Following the above logic, and for the convenience of the data interpretation, Table 2 provides individual ranks for each affinity prediction method according to each of three assessment techniques, as well as a consensual performance measure expressed as an average of the three individual ranks. In the next sections, we propose a detailed comparative analysis of the results obtained and computational protocols used.

MM-PB/SA outperforms MM-GB/SA

All dynamics-based predictions including MM-PB/SA and MM-GB/SA were obtained by processing snapshots from two independent 20 ns MD simulations. The gas phase interaction term ΔE_{MM}, polar solvation term ΔG_{solv} approximated by PB and GB equations, as well as the non-polar solvation term ΔG_{SA} are summarized in Supplemental Table 1. The binding energies predicted by MM-PB/SA are in a good agreement with experimental data with an R^2 value of 0.88 (Table 2 and Figure 1). It also effectively discriminates among the three apparent groups of weak (**1**, **4**), moderate (**2**, **3**, **5**) and stronger (**6**, **7**) binders (as can

be seen from Figure 1). The only minor inaccuracy is an inversion of the predicted rank orders for compounds **6** and **7**. However, on the energetic scale, MM-PB/SA has predicted quite unrealistic figures. For example, the predicted affinities of the compounds **2**, **3**, **5**, **6** and **7** lie in the range of ~5.2 kcal/mol (from −31.98 to −26.79 kcal/mol), which is nearly 3.5-fold as compared to the experimentally determined range of ~1.5 kcal/mol (from −5.59 to −7.06 kcal/mol). This large energetic range was also observed when MM-PB/SA is applied in several other protein systems[41,42]. Hence, unlike FEP, MM-PB/SA can only be considered as a ranking tool rather than a true free energy predictor in the case of the KMe reader L3MBTL1. MM-PB/SA is ranked 2nd according to R^2 (0.88), 5th according to τ_b (0.68) and 3rd according to PI (0.95) that, with an average rank of 3.3, places it third in the consensus ranking.

The dynamics-based MM-GB/SA method has shown a significantly lower performance than MM-PB/SA as instantiated by all three correlation coefficients. Moreover, MM-GB/SA was not able to discriminate between the weak, moderate and strong binders. For example, despite a large predicted energetic scale, the computed free energy difference between a moderate binder **3** (−28.7 kcal/mol) and a non-binder **4** (−28.0 kcal/mol) was just 0.7 kcal/mol. The only bright spot in the poor outcome of MM-GB/SA predictions was a robust discrimination between the weakest and the strongest binders. It should be noted that the mediocre performance of MM-GB/SA is, to some extent, in contrast with the results of a recent study by Hou *et al*[43]. In their comparative analysis performed on a set of six proteins (structurally different from L3MBTL1), MM-GB/SA was a slightly better predictor than MM-PB/SA on most protein targets.

MM-PB/SA tolerates short simulation times

We then investigated on how to optimize the computational efficiency of the MM-PB/SA method. First, we made predictions by means of a single-point version of MM-PB/SA, in which only a local energy minimization on docking poses was conducted instead of the MD conformational sampling. In recent years, this method has become a popular rescoring scheme for the docking-based virtual screening and, according to some reports, demonstrated a better predictive power as compared to scoring functions embedded in the docking software[44–47]. In the current study, the single-point MM-PB/SA was significantly less predictive than its dynamics-based counterpart (Table 2). Therefore, the next step was to determine the optimal simulation time. We made use of our 20 ns simulations in order to calculate MM-PB/SA free energy values at progressively increasing simulation times starting from 0.1 ns. The plot in Figure 2 demonstrates that even relatively short simulations lead to high quality predictions. For instance, the shortest tested simulation results in a best possible compound ranking (PI = 0.95). Additional 3 ns are then necessary to obtain the highest possible R^2.

It is worth to note that a too short simulation may suffer from statistical uncertainty. This is especially important for MM-PB/SA and MM-GB/SA as these calculations sometimes yield a large standard deviation of the estimated free energy[48]. As pointed out by a recent article[48], 1.5 to 15 ns simulation is needed to obtain a statistical precision of 1 kJ/mol (0.24 kcal/mol) for seven complexes of avidin-biotin analogues. This is nearly an order of magnitude longer than the length of most reported MM-PB/SA simulations. In our study, the standard deviations of the binding free energies obtained from alternative snapshot samplings varied from 3 to 5 kcal/mol (Supplemental Table 1), which is quite low compared to a typical reported values of 5 to 35 kcal/mol[48]. Hence, at the sampling frequency of 10 ps, a 3 ns simulation was sufficient to obtain a standard error of mean (SEM) around 0.25 kcal/mol. Finally, simulation times longer than 3 ns did not result in any increase of the prediction quality in this study.

In all simulations the actual data have been collected after a short equilibration time. For obvious reasons, this equilibration step is especially important for short simulations in order to "smoothen" unfavorable interactions mainly due to soaking the protein-ligand system in the explicit solvent medium. In our setting, a 0.3 ns equilibration was sufficient for a convergent "production" run. However, certain compounds required a longer time. For example, compound **4** needed at least a 1 ns equilibration, mainly because the bulky tri-methyl amino group did not fit the initial receptor structure. However, given that compound **4** is a very weak binder, even a partially equilibrated structure did not dramatically affect the prediction quality.

Deficient performance of LIE

The LIE method was applied in four different settings including the dynamics-based and single-point protocol, each with either predetermined α and β parameters or fitted to our data in order to reach a better correlation. None of these protocols has demonstrated an adequate performance as reflected by all three statistical criteria (Table 2). When using the recommended parameter set (α = 0.18 and β = 0.50), predicted binding affinities correlated very weakly with experimental data (R^2 = 0.16, τ_b = 0.43, PI = 0.42). We have also tried to fit the scaling factors for a better match with experimental binding energies. However, the fitted parameter set (with α = 0.71 and β = 0.17) only led to an insignificant gain in R^2 to 0.29. In terms of compound ranking, LIE method can clearly distinguish two strong binders **6** and **7** from the other five compounds. On the other hand, the binding affinity of compounds **1** and **4** was overestimated, partially due to an unexpected large electrostatic term. In addition, LIE incorrectly predicted that there is a large free energy difference among the three weak binders **2**, **3** and **5**.

Here we used our previous free energy decomposition results[9] to analyze the possible deficiencies of LIE terms in this test. Although there is no direct one-to-one correspondence, a qualitative relationship can be anticipated between the polar and non-polar FEP terms and the two energy terms in LIE. For instance, our previous decomposition suggested that compound **2** benefits from more efficient vdW interactions and higher bound-state mobility and that compounds **3**, **5** and **6** have a higher electrostatic contribution due to their highly localized positive charge on the hydrogen pointing to Asp355. In LIE, the relative values of the vdW term $\langle \Delta E_{l-s}^{vdW} \rangle$ are clearly correlated with our FEP non-polar contributions (Supplemental Table 2). On the other hand, the electrostatic term $\langle \Delta E_{l-s}^{elec} \rangle$ is practically uncorrelated to the electrostatic contributions from our FEP study. Moreover, it has already been reported that LIE may have issues in treating charged ligands, particularly their electrostatic terms[37,49]. A possible solution might involve constant pH simulation or a polarizable force field. Another LIE problem in this study was an excessively high variation of both energy terms of Equation 2 across different MD snapshots. The latter fact is in contrast with a relative stability of the MM-PB/SA energy terms, as well as with a good convergence of our FEP simulations on the same molecular systems[9].

In addition to the original LIE scheme, we also examined the single-point LIE method. This scheme was performed on a single minimized structure with implicit SGB solvation. Using the default scaling factors, a very pour correlation is found (R^2 = 0.34). After fitting the scaling factors to experimental data, R^2 increased to a still low 0.49.

Scoring functions

A total of 10 scoring functions were evaluated in this test, including GlideSP, GlideXP from Schrodinger[21]; Ligscore[19], PLP[18], Jain[20], PMF[16] and PMF04[17], as implemented in Discovery Studio; Dock6 grid score and contact score in Dock6[14]; as well as AutoDock4

score[15]. These 10 scoring functions represent all three major types of energy functions, i.e. force-field-based, knowledge-based and empirical-based. Before scoring each ligand, we performed a local minimization to optimize the initial binding pose. In Discovery Studio, a minimization step by means of the CHARMM force field was performed prior to the scoring step. In Maestro, Dock6 and AutoDock4, we relied upon native energy minimization methods intrinsic to the scoring algorithms.

As can be seen from Table 1, the 10 scoring methods performed quite differently in this study, with R^2 ranging from 0.28 to 0.95. Three scoring functions (PLP, PMF04 and Autodock4) demonstrated a very high R^2 value (>0.70). There are three scoring functions (GlideSP, LigScore and Contact) where the correlation is almost negligible ($R^2 < 0.40$). In general, the rank-order correlation coefficients τ_b and PI followed the same trend as that of R^2, which ranks both PLP and PMF04 as the top two performers. However, there are a few exceptions: for example, GlideSP has a poor R^2 value of 0.39, mainly because of a single outlier (compound **7**), while it has correctly ranked the other six molecules that resulted in a PI value of 0.87. Remarkably, the best scorer, PMF04, is also the best overall performer in the current study, ahead of computationally demanding MM-PB/SA. However, a larger number of compounds would be necessary to confirm this finding.

Interestingly, the top scorers represent all three major categories of scoring functions. Piecewise Linear Potential (PLP) is an empirical function that accounts for all atom-atom interactions intermolecular interactions[18]. Potential of Mean Force 04 (PMF04) is a knowledge-based function derived from statistical analysis of protein ligand complexes[16,17]. It outperformed its ancestor, PMF, probably due to a larger parameterization data set. Autodock4 is a semi-empirical energy function, which combines force field empirical weights and/or empirical functional forms[15]. Finally, a force-field-based function, Dock6, performed moderately well in this test ($R^2 = 0.62$). Dock6 score is a grid-based score derived from the non-bonded terms of a molecular mechanics force field[14].

Six of the ten scoring functions have correctly predicted compound **7** to have the strongest binding affinity in this set. Among the six functions, only PLP and PMF04 have correctly ranked compound **6** to be the 2nd strongest binder. Another challenge was to correctly rank the binding affinities of moderate binders **2**, **3** and **5**, whose experimental binding free energies differ only by ~0.2 kcal/mol. Of all scoring functions, only PLP, PMF04 and GildeSP have correctly predicted these three compounds to have comparable bind affinities. On the other hand, most scoring functions were correct in ranking compounds **1** and **4** as the weakest binders. This is likely due to the ability to recognize a weaker electrostatic interaction with Asp355 that was experimentally demonstrated to be critical for ligand binding[6].

Interestingly, the average R^2 over 10 scoring functions amounts to 0.56 that, despite the novelty of the protein target, is overall superior than what was observed in earlier benchmark studies on more familiar therapeutic targets[22,23]. For instance, in a test against 800 diverse protein-ligand complexes, the R^2 values demonstrated by 14 energy functions ranged from 0.02 to 0.32[22]. The authors have also evaluated the performance of those scoring functions on three individual targets: HIV-1 protease, trypsin and carbonic anhydrase II and found that most scoring functions failed to correctly rank ligands. The R^2 values ranged from 0.04 to 0.41, with an average of 0.20[22]. Another study that involved eight receptors and ten docking programs resulted in R^2 values ranging between 0.01 and 0.16[23]. The authors concluded that "no single program performed well for all of the targets and none of the docking programs made a useful prediction of the ligand binding affinity". In our study, the surprisingly good performance of the standard scoring functions may have two complementary explanations: (i) scoring functions are subject to continuous

improvements (e.g., superior performance of PMF04 as compared to the original PMF), (ii) the MBT binding pocket is relatively small and forms fewer interactions with the ligand as compared to other proteins that leads to a smaller cumulative error.

Conclusions

In this study, we evaluated how various computational predictors of the ligand-protein binding affinity perform on a series of seven novel small-molecule binders to a methyllysine histone code "reader", L3MBTL1. As a consensus of the comparison metrics used, the study has revealed three top performers, far ahead of the other techniques. They include two scoring functions, PMF04 ($R^2 = 0.95$; $\tau_b = 1.0$; PI = 1.0) and PLP ($R^2 = 0.80$; $\tau_b = 0.81$; PI = 0.98); and a simulation-based method MM-PB/SA ($R^2 = 0.88$; $\tau_b = 0.68$; PI = 0.95). It should be noted that while PMF04 and PLP show excellent results on L3MBTL1, their performance varies greatly from one protein target to another[22,23]. Therefore, their general utility for the KMe histone code readers is still to be confirmed when more experimental data become available. Alternatively, MM-PB/SA is a less empirical and more CPU-intensive approach (compared to PMF04 and PLP) with a thorough handling of solvation effects. Therefore, despite being ranked only third on the current data set, it is more likely to demonstrate a comparable predictivity on other ligands and proteins similar to those used in the current analysis. Here, we report a computational protocol that enables an optimal speed-predictivity ratio when applying MM-PB/SA to histone KMe readers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Rodriguez-Paredes M, Esteller M. Nat Med. 2011:330–339. [PubMed: 21386836]

2. Turner BM. Nature Cell Biol. 2007; 9:2–6. [PubMed: 17199124]

3. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Cell. 2010; 143(3):470–484. [PubMed: 21029866]

4. Bonasio R, Lecona E, Reinberg D. Semin Cell Dev Biol. 2009; 21(2):221–230. [PubMed: 19778625]

5. Trojer P, Reinberg D. Cell Cycle. 2008; 7(5):578–585. [PubMed: 18256536]

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. 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(31):4293–4304. [PubMed: 18408754]

8. Kireev D, Wigle TJ, Norris-Drouin J, Herold JM, Janzen WP, Frye SV. J Med Chem. 2010; 53(21):7625–7631. [PubMed: 20931980]

9. Gao C, Herold JM, Kireev DB, Wigle TJ, Norris JL, Frye SV. J Am Chem Soc. 2011; 133(14):5357–5362. [PubMed: 21428286]

10. 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; 54(7):2504–2511. [PubMed: 21417280]

11. Jhoti, H.; Leach, A. Structure-based Drug Discovery. Springer; 2007.

12. Adams-Cioaba MA, Min J. Biochemistry and Cell Biology. 2009; 87(1):93–105. [PubMed: 19234526]

13. 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(12):1229–1230. [PubMed: 18026117]

14. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. RNA. 2009; 15(6):1219–1230. [PubMed: 19369428]

15. Huey R, Morris GM, Olson AJ, Goodsell DS. J Comput Chem. 2007; 28(6):1145–1152. [PubMed: 17274016]

16. Muegge I, Martin YC. J Med Chem. 1999; 42(5):791–804. [PubMed: 10072678]

17. Muegge I. J Med Chem. 2005; 49(20):5895–5902. [PubMed: 17004705]

18. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DR, Fogel LJ, Freer ST. Chem Biol. 1995; 2(5):317–324. [PubMed: 9383433]

19. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM, Waldman M. J Mol Graphics Model. 2005; 23(5):395–407.

20. Jain AN. J Comput-Aided Mol Des. 1996; 10(5):427–440. [PubMed: 8951652]

21. Schrödinger LLC. 2010

22. Wang R, Lu Y, Fang X, Wang S. J Chem Inf Comp Sci. 2004; 44(6):2114–2125.

23. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. Journal of Medicinal Chemistry. 2006; 49(20):5912–5931. [PubMed: 17004707]

24. Bowers, KJ.; Chow, E.; Xu, H.; Dror, RO.; Eastwood, MP.; Gregersen, BA.; Klepeis, JL.; Kolossváry, I.; Moraes, MA.; Sacerdoti, FD.; Salmon, JK.; Shan, Y.; Shaw, DE. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06); Tampa, Florida. 2006.

25. Shaw, DE. Research. New York, NY: 2009.

26. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. J Chem Phys. 1995; 103:8577.

27. Shan Y, Klepeis JL, Eastwood MP, Dror RO, Shaw DE. J Chem Phys. 2005; 122:54101. [PubMed: 15740304]

28. Ryckaert J, Ciccotti G, Berendsen H. Journal of Computational Physics. 1977; 23:327–341.

29. Andersen H. Journal of Computational Physics. 1983; 52:24–34.

30. Pastor R, Brooks B, Szabo A. Mol Phys. 1988; 65:1409–1419.

31. Feller SE, Zhang Y, Pastor RW, Brooks BR. J Chem Phys. 1995; 103:4613.

32. Gohlke H, Case DA. J Comput Chem. 2004; 25(2):238–250. [PubMed: 14648622]

33. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Crowley, M.; Walker, RC.; Zhang, W.; Merz, KM.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Kollman, PA. University of California; San Francisco: 2008.

34. Grant JA, Pickup BT, Nicholls A. J Comput Chem. 2001; 22(6):608–640.

35. Åqvist J, Medina C, Samuelsson JE. Protein Eng. 1994; 7:385–391. [PubMed: 8177887]

36. Wang W, Wang J, Kollman PA. Proteins. 1999; 34:395–402. [PubMed: 10024025]

37. Åqvist J, Marelius J. Combinatorial Chem High Throughput Screening. 2001; 4(8):613.

38. Carlsson J, Andér M, Nervall M, Åqvist J. J Phys Chem B. 2006; 110:12034–12041. [PubMed: 16800513]

39. Huang D, Caflisch A. J Med Chem. 2004; 47:5791–5797. [PubMed: 15509178]

40. Charifson PS. J Med Chem. 2001; 44(21):3417–3423. [PubMed: 11585447]

41. Brown SP, Muchmore SW. J Chem Inf Model. 2006; 46(3):999–1005. [PubMed: 16711718]

42. Guimarães CRW, Mathiowetz AM. J Chem Inf Model. 2010; 50(4):547–559. [PubMed: 20235592]

43. Hou T, Wang J, Li Y, Wang W. J Chem Inf Model. 2011; 51(1):69–82. [PubMed: 21117705]

44. Lyne PD, Lamb ML, Saeh JC. Journal of Medicinal Chemistry. 2006; 49(16):4805–4808. [PubMed: 16884290]

45. Guimarães CRW, Cardozo M. J Chem Inf Model. 2008; 48:958–970. [PubMed: 18422307]

46. Thompson DC, Humblet C, Joseph-McCarthy D. J Chem Inf Model. 2008; 48:1081–1091. [PubMed: 18465849]

47. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M. J Med Chem. 2005; 48:4040–4048. [PubMed: 15943477]

48. Genheden S, Ryde U. J Comput Chem. 2010; 31(4):837–846. [PubMed: 19598265]

49. Åqvist J. J Comput Chem. 1996; 17(14):1587–1597.
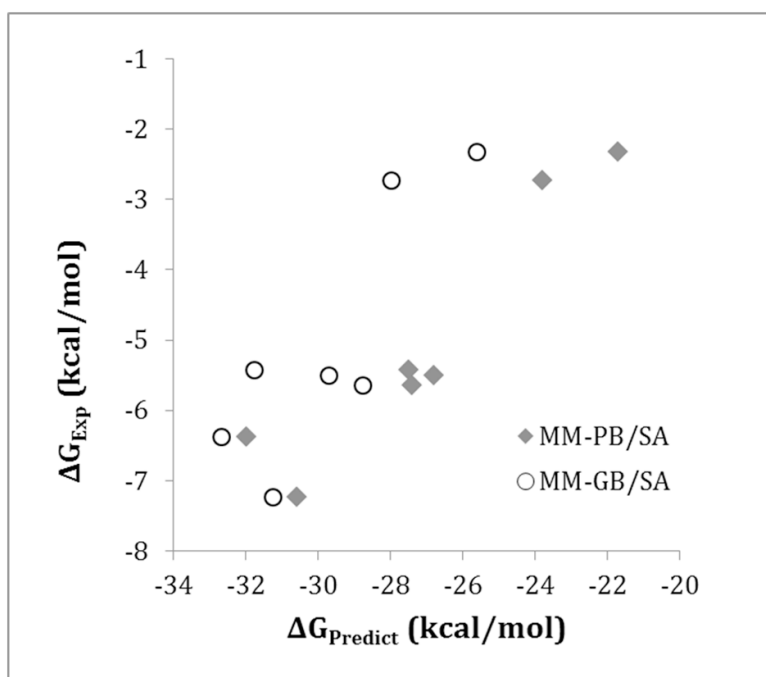
Figure 1.
Scatter plots of predicted MM-PB/SA (diamonds) and MM-GB/SA (circles) binding free energies versus experimental data.
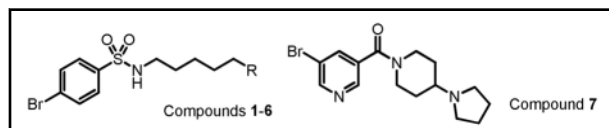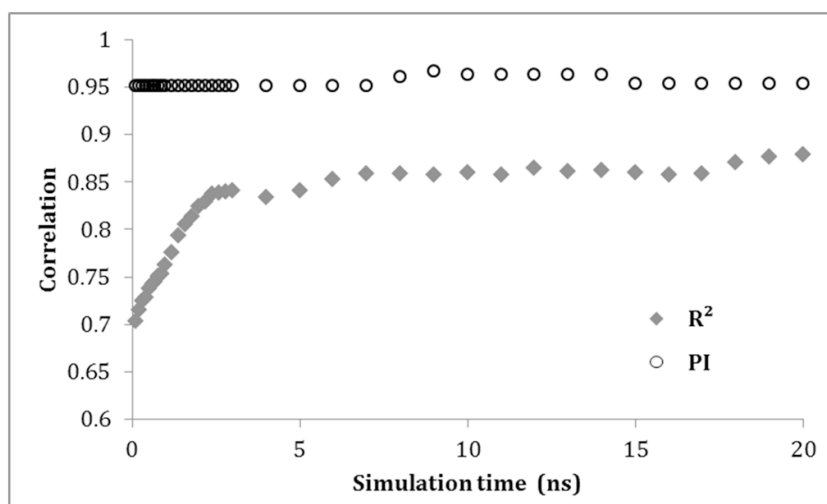
Figure 2.
Impact of simulation time on the predictive power of MM-PB/SA represented by R^2 (circles) and PI (diamonds).

Table 1

Structures of seven designed biophysical probes and their experimental binding affinities to L3MBTL1

Compounds 1-6   Compound 7

N	Name	R	K_d (μM)	ΔG_{Exp}
1	UNC587	N	>20000*	>−2.32*
2	UNC588	NMe	104 ± 10	−5.43
3	UNC589	NMe$_2$	73 ± 19	−5.64
4	UNC590	NMe$_3$	>10000*	>−2.73*
5	UNC592	N(Me)Et	91 ± 11	−5.51
6	UNC591	Pyrrolidine	21 ± 2	−6.38
7	UNC669		5 ± 1	−7.23

Dissociation constants K_d are measured by ITC and are the average of two independent measurements[9, 10]. The ITC experiments were performed at 25°C. Free energy values (ΔG_{Exp}) are expressed in kcal/mol.

*: Neither of compounds **1** and **4** showed any binding interaction in ITC experiment. Hence, their K_d values are estimated from the experimental detection limit. The relative binding affinities of **1** and **4** were extrapolated from data by Li *et al*[6].

Table 2

The summary of computational results for all tested free energy predictors.

Method category	Method	Individual ΔG$_{bind}$ predictions for compounds 1–7							Overall method performance			
		1	2	3	4	5	6	7	R² (rank)	τ$_b$ (rank)	PI (rank)	⟨Rank⟩
Experiment (kcal/mol)	ITC	>−2.32	−5.43	−5.64	>−2.73	−5.51	−6.38	−7.23	1.00	1.00	1.00	
	FEP*	4.73	0.00	−0.30	1.81	−0.65	−1.82	N/A	0.88	0.87	0.99	
Simulation-based Methods (kcal/mol)	PB/SA	−21.69	−27.50	−27.40	−23.78	−26.79	−31.98	−30.60	0.88 (2)	0.68 (5)	0.95 (3)	3.3
	GB/SA	−25.62	−31.77	−28.77	−27.97	−29.71	−32.69	−31.26	0.71 (5)	0.49 (11)	0.86 (6)	7.3
	LIE	−6.54	−1.85	−3.57	−3.85	−7.74	−8.38	−9.78	0.16 (18)	0.43 (16)	0.42 (18)	17.3
	LIE fitted	−4.16	−4.46	−3.83	−5.03	−4.79	−5.96	−6.52	0.30 (16)	0.43 (15)	0.45 (17)	16
Single-point Rescoring (kcal/mol)	PB/SA	−35.39	−41.67	−43.85	−38.94	−46.64	−43.29	−43.83	0.69 (6)	0.49 (12)	0.77 (10)	9.3
	GB/SA	−37.14	−44.42	−42.17	−43.06	−41.69	−51.72	−43.45	0.34 (14)	0.39 (17)	0.59 (13)	14.7
	LIE	−12.21	−14.89	−18.39	−16.59	−17.35	−18.32	−16.12	0.34 (13)	0.43 (13)	0.47 (15)	13.7
	LIE fitted	−3.17	−4.58	−5.64	−4.14	−5.21	−7.24	−4.78	0.49 (9)	0.71 (4)	0.83 (8)	7
Scoring functions	GlideSP	−5.07	−5.20	−5.28	−4.46	−5.42	−6.62	−5.34	0.39 (12)	0.62 (8)	0.87 (4)	8
	GlideXP	−7.63	−8.19	−7.56	−6.38	−8.72	−8.24	−8.05	0.42 (11)	0.24 (18)	0.59 (14)	14.3
	LigScore	−2.15	−1.96	−2.73	−1.33	−3.04	−3.26	−2.17	0.31 (15)	0.43 (14)	0.66 (12)	13.7
	PLP	−51.89	−55.31	−65.23	−45.94	−66.66	−67.76	−69.71	0.80 (3)	0.81 (2)	0.98 (2)	2.3
	Jain	−2.22	−1.46	−3.52	−2.56	−5.64	−5.03	−6.62	0.49 (10)	0.62 (7)	0.69 (11)	9.3
	PMF	−67.68	−72.10	−83.17	−74.39	−86.01	−77.01	−96.27	0.56 (8)	0.62 (6)	0.80 (9)	7.7
	PMF04	−46.72	−56.45	−59.32	−49.27	−57.27	−59.97	−67.43	0.95 (1)	1.00 (1)	1.00 (1)	1
	Autodock4	−3.38	−3.95	−4.56	−4.00	−4.99	−4.76	−5.22	0.74 (4)	0.71 (3)	0.83 (7)	4.7
	Dock6	−33.38	−36.35	−36.34	−30.04	−40.06	−42.20	−37.50	0.62 (7)	0.52 (10)	0.86 (5)	7.3
	Contact	−115	−123	−143	−147	−132	−145	−156	0.28 (17)	0.62 (9)	0.45 (16)	14

*: For FEP method, computed binding free energies are given relative to the binding free energy of compound **2**, which were obtained in our previous study[9]. The performance of FEP was evaluated using compounds **1**–**6** only.