Conformational and Thermodynamic Differences Underlying Wild-type and Mutant YEATS Domain Specificity for Epigenetic Marks

Lokesh Baweja[†,‡] and Jeff Wereszczynski[*,†,‡]

†*Department of Physics, Illinois Institute of Technology, Chicago, USA*

‡*Center for Molecular Study of Condensed Soft Matter, Illinois Institute of Technology, Chicago, USA*

E-mail: jwereszc@iit.edu

Abstract

Histone post-translational modifications (PTMs) are interpreted by multiple reader domains and proteins to regulate gene expression. The YEATS domain is a prototypical PTM reader that recognizes multiple lysine acetylation marks on the histone H3 tails as a way of recruiting chromatin remodellers. Two YEATS mutations have been identified which have been linked with leukemia, Wilms tumor, and other forms of cancer and result in either an insertion or deletion of residues in the loop connecting beta sheets distant from the protein active site. *In vitro* experiments have shown that these mutations modulate the selectivity of YEATS domains for various lysine acetylation marks, although different experiments have provided contrasting views on the ability of the insertion and deletion mutants to discern specific PTMs. Here, we have performed multiple molecular dynamics simulations of wild-type, insertion, and deletion mutant YEATS domains free from and in complex with two PTM mimicking peptides: one that mimics acetylation on residue K9 of H3 and the other that mimics

acetylation on residue K27 of H3. Results show that these two peptides have distinct flexibilities and binding energetics when bound to YEATS domains, and that these properties are affected by interactions with residues within and outside of the peptide consensus motif. Furthermore, these properties are modulated by the YEATS insertion and deletion mutants, which results in disparate binding affects in these systems. Together, these results suggest that only the partial exposure of histone tails is sufficient in the context of nucleosomes for YEATS-mediated recognition of acetylation marks on histone tails. They also caution against the over-interpretation of results obtained from experiments on reader domain-histone peptide binding in isolation and not in the full-length nucleosome context.

Introduction

Chromatin reader proteins process epigenetic information by recognizing various posttranslational modifications (PTMs) on the histone tails.[1] Eleven nineteen leukemia (ENL) binds with acetylated chromatin regions through its YEATS domain (named for the first discovered members of the family: Yaf9, ENL, AF9, Taf14, Sas5)[2,3] and recruits super elongation complex that regulates transcription of ENL target genes through RNA polymerase II.[2,4] Oncogenic transformations in acute myeloid leukaemia are linked with acylation reading of ENL, and inhibition of YEATS domain via small molecules is suggested as a promising therapy to treat leukaemia.[5–8] YEATS adopts an immunoglobulin-like topology with an elongated β- sandwich core with four loops (L1, L4, L6, and L8) connecting the β-strands. The narrow hydrophobic binding site is formed by the three loops (L1, L4, and L6) and lined by histidine, serine, and tyrosine.[9] YEATS binds to Nε-acetylated lysine (Kac) in histone tails at positions K9 and K27, and in turn, stabilizes the transcription machinery for the expression of target genes.[3,10,11] To date, eight mutations have been discovered in ENL protein causing the insertion and deletion in the L8 region of the YEATS domain.[12,13] Among the known ENL mutants, insertion (p.117_118insNHL) and deletion (p.111_113NPP>K) mutations have been found at a high frequency within diseased patients and nearly 5 percent of people having Wilms tumor.[12,14]

In vitro experiments performed using isolated histone peptide mimics (H3K9ac and H3K27ac) have shown that these insertion and deletions mutants bind H3K9ac and H3K27ac peptides with varied thermodynamic affinities.[12,15] Moreover, some experiments have shown contrasting results in the affinities of these mutant with H3 peptide mimics. For example, Perlman *et al.*, showed that the insertion mutation abolished the interaction between YEATS and H3K27ac,[12] whereas Wan *et al.*, reported a slightly lower binding affinity of the insertion mutant with H3K27ac than wild type YEATS.[15] Moreover, recent findings using isothermal calorimetry (ITC) experiments reported similar binding affinities of wild-type and the deletion mutant for H3K9ac but with different thermodynamic mechanisms.[16] Interestingly,

structural studies have shown that the overall structure of YEATS remains largely consistent across mutations, despite the lengthening and shortening of the L8 loop due upon insertion or deletion.[16] Therefore, the mechanisms underlying these disparate binding affinities and how they are affected by mutations remains an open question.

To resolve these differing experimental results and understand the molecular level effects of mutations on the binding of YEATS to acetylated histones, we performed multiple microsecond long molecular dynamics (MD) simulations. Here, we have focused on how conformational and thermodynamic differences induced by mutations affect the binding of H3K9ac and H3K27ac peptides. MD simulations revealed that the binding of H3K9ac with YEATS is thermodynamically more favorable than H3K27ac. Residues beyond the acylation recognition signature (−ARKacS−) in the peptide mimics interacted dynamically but did not contribute significantly to binding with the YEATS domains. Hence, in the full-length nucleosome, only partial exposure of residues is likely required for recognition of acetylation marks on histone tails by YEATS. We also noted that H3K9ac binding highly restricts the conformational space of the deletion mutant in comparison to the wild-type and insertion mutant. Simulations also show robust binding of H3K9ac peptide to the insertion mutant, in contrast to some previous experimental studies which show no binding.[12,16] In addition, in the context of the nucleosome, the accessibility of histone-acetylation sites (K9, K27) and other positively charged residues is masked by their strong interactions with DNA.[17–19] Therefore, the specifcity of epigenetic modifications is also likely controlled by higher order interactions of histone tails with nucleosomal DNA, which is not captured by our simulation study and previous experiments.

Materials and Methods

Structure Models

Structure of wild-type YEATS domain was obtained from the 5J9S,[11] deletion and insertion mutant from PDBs 7B10[16] and 7B0T,[16] respectively for apo simulations (Figure 1).

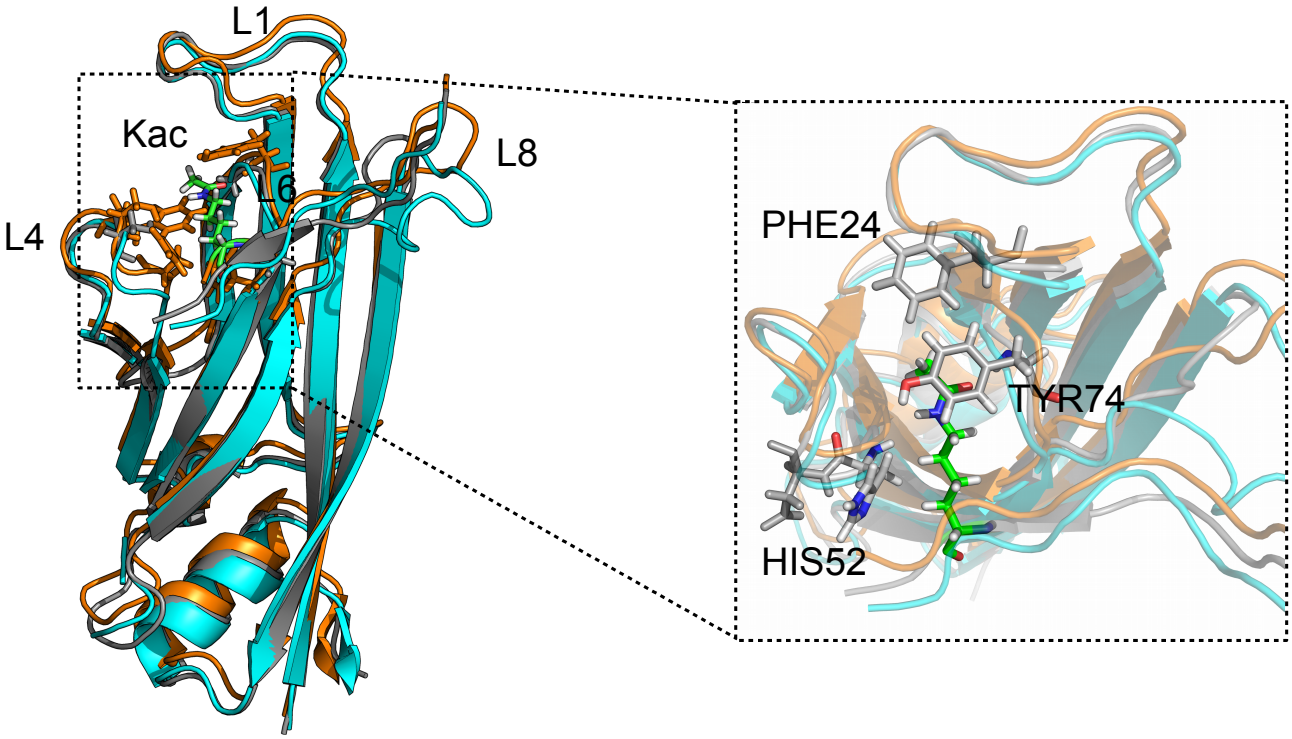

Figure 1: Cartoon representation of wild-type and deletion and insertion mutant of ENL YEATS domain in complex with full-length H3K9ac peptide.Wild-type YEATS domain is shown with orange, deletion mutant in gray and insertion mutant in cyan color. Zoomed-in view of Kac pocket is provided with the active residues displayed in stick representation.

Peptide bound state of wild-type YEATS domain from 5J9S[11] was used for preparing initial structures for simulations by increasing the length of the peptide using MODELLER[20] to mimic the 11 residue peptide used for isothermal calorimetry experiments in previous studies.[12,21] Further, structures from 7B10[16] and 7B0T[11] were aligned with 5J9S to generate bound states of insertion and deletion mutant of YEATS. The peptide mimic sequences used for H3K9ac and H3K27ac were (1-11H3)ARTKQTAR(Kac)ST and (19-29H3)QLATKAAR(Kac)SA, respectively. Both the peptides share a common −ARKacS−

motif. We divided the peptide sequences into two regions as consensus (−ARKacS−) and non-consensus (H3K9ac(ARTKQT) and H3K27ac(QLATKA)) motifs.

MD simulations

All simulations were conducted using the AMBER software suite with CUDA-enable PMEMD engine of version (v18).[22,23] The Amber 14SB force-field parameters along with OPC water models were used in this study.[24,25] The force-field parameters for acetylated lysine were obtained from the previous study.[26] A 4 fs time-step was used in conjunction with SHAKE and hydrogen mass repartitioning for all simulations.[27,28] All systems were first energy minimized for 5000 steps with a solute harmonic restraint of 10 kcal/mol/$Å^2$, followed by 5000 steps without restraints. Equilibration was performed in two steps including 100 ps of constant volume simulations where the temperature was gradually heated from 10 to 310 K. Then, the heavy atom restraints were gradually released over 500 ps of NPT run. The pressure was controlled via a Monte Carlo barostat with a target pressure of 1 atm and a relaxation time of 3.0 ps^{-1}.[29] Production runs were performed at 310 K using a Langevin thermostat.[30,31] We performed ten,[32,33] 1μs simulations per system in the NPT ensemble, accumulating 90μs of sampling across all the three systems. Trajectories were recorded every 10 ps and visualized using VMD[34] and PyMol.[35] Trajectories analyses were performed on the last 750 ns of the simulations, allowing for 250 ns of equilibration using CPPTRAJ[36] and in-house python scripts.

Simulation analyses

Root mean square fluctuation (RMSF) analysis was performed on the apo and peptide bound (H3K9ac and H3K27ac) states of wild-type and mutant YEATS domain. For RMSF calculations, trajectories were aligned to their initial configuration. Translations and rotations were removed by least-squares fitting the backbone of the YEATS domains, and RMSFs were computed on the Cα atoms. Reported RMSFs are the average of all 10 simulations with

the standard error of the mean obtained from 10 samples for each system. Kullback-Leibler Divergence was performed using internal coordinates (backbone and side-chain dihedrals) to compare conformational ensembles across the systems.[37,38] Contact analysis between YEATS and histone peptides was performed using MDanalysis[39] and in-house python scripts, where contacts were defined as between heavy atoms of histone peptides residues that were within a distance of 4.5 Å from YEATS heavy atoms. Interaction energies between the YEATS and acetylated peptides were determined via an MM-GBSA (Molecular Mechanics Generalized Born Surface Area) analysis with igb=8 and a salt concentration of 0.15 M.[40,41] Error bars represent the standard error of the mean, with a decorrelation time of 20 ns that is based on a statistical inefficiency test of MM/GBSA values.[42]

Principal component analysis (PCA)

PCA is used to extract slow and functionally important motions of YEATS domains in apo and peptide bound states from the MD trajectories. First, the covariance matrix of atomic positions of Cα atoms of wild-type and mutant YEATS domain was built, and the eigenvectors of this matrix (also known as principal components) describes the concerted motion of the system.[43] In this study, PCA was performed using GROMACS-2016.3[44] to investigate the similarities and differences in the conformational ensembles of wild-type and mutant YEATS domain.

Mutual Information based Generalized Correlation Analysis

Generalized correlation analysis relies on mutual information theory.[30,45] Cα atoms of YEATS domain present at two distant sites are correlated if their joint probability distribution of fluctuation is less than the product their marginal probability distribution. This method captures the linear as well as non-linear correlations but cannot distinguish if the two sites are positively or negatively correlated.

Results

Mutations and Peptide Binding Alter Loop Dynamics

To assess how mutations and substrate binding affect YEATS domain dynamics, a series of MD simulations were performed. For the wild type, insertion, and deletion mutations, 10 independent 1 µs simulations were run performed with no substrate bound. Additional sets of 10x1 µs simulations were performed for systems with H3K9ac and H3K27ac peptide mimics bound for each protein to determine the disparate effects these two substrates have on system structures, dynamics, and thermodynamics. In each simulation the YEATS domain maintained a well-folded state with minimal overall structural changes (Figure S1). An analysis of protein fluctuations pinpointed two regions of the YEATS domain in which the dynamics were altered by mutations or peptide binding: loop 1 and loop 8. Root-mean-square fluctuation (RMSF) analysis of protein Cα atoms show that the majority of each protein has minimal dynamics, with RMSF values of 1.0-1.5 Å (Figure 2). In loop-8 there were significant differences in RMSF values between the wild-type, deletion, and insertion mutants, regardless of the peptide binding state. In the wild-type, fluctuations were on the order of 2.0 Å (Figure 2a). These fluctuations were reduced in the deletion mutant to ∼1.5 Å, but they were increased to ∼4.0 Å in the insertion mutant (Figure 2b, 2c). This dramatic increase is due to the increased length of this loop in the insertion mutant, whereas in the deletion mutant the loop is shorter which restricts its range of motion.

To probe the effects of mutations and peptide binding on YEATS domain internal dynamics, the distributions of backbone and side-chain dihedral motions were compared between systems with Kullback-Leibler divergence (KLD) calculations. In the apo state, these calculations show that the dynamics are most perturbed by the deletion mutation, which created high KLD values throughout the protein relative to both the wild-type and insertion systems (see Figure S3). Meanwhile, the KLD values upon substrate binding were highest upon H3K9ac binding to the wild-type and insertion mutants (Figure S2).

Mutational studies showed that histone peptide recognition is mainly dependent on the Tyr74 located above the acetylated lysine binding pocket.[11] To monitor the conformation of Tyr74, the distributions of the two sidechain dihedrals, $\chi1$ and $\chi2$, were computed. In the apo wild-type YEATS, Tyr74 exists in a dynamic equilibrium between stacked and unstacked states, with $\chi1$ sampling four states and $\chi2$ independently sampling two, for a total of eight energy wells (Figure 3). Upon peptide binding there was a dramatic shift in these distributions. Binding of the H3K27ac peptide largely restricted $\chi1$ to the well around 100°), with similar sampling of $\chi2$. In contrast, the binding of H3K9ac to the wild-type shifts the distribution of $\chi2$ towards the positive values, where Tyr74 hydroxl group is lying anti-parallel to the peptide's acetylated lysine. The Tyr74 conformations were largely unaffected by the insertion mutant, with the apo conformation sampling multiple energy wells which were restricted by peptide binding (Figure 3c). However, the deletion mutant created a dramatic shift in the distributions of the $\chi1$ and $\chi2$ angles. In particular, the $\chi1$ angles were primarily centered around $\pm180°$ in the apo state, which was maintained in the H3K27ac bound state. In contrast, upon binding H3K9ac, $\chi1$ shifts to ~0 ° and $\chi2$ is restricted to $\sim$-100°.

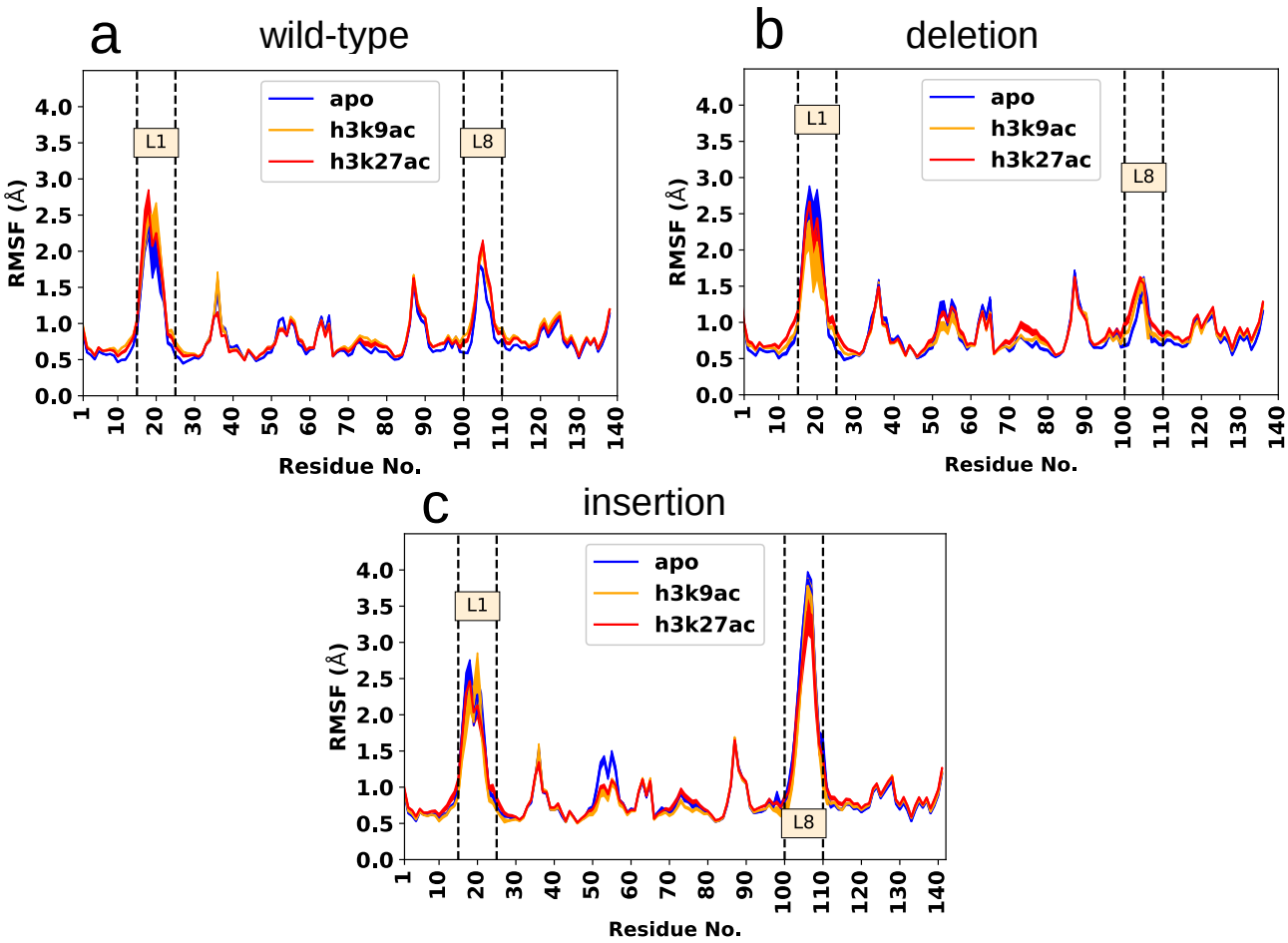

Figure 2: Root mean square fluctuation of Cα atoms present in wild-type and mutant yeats domain with respect to their initial conformation analyzed over 10 replicates. a) wild-type, b) deletion, and c) insertion

MM/GBSA analyses were used to estimate the binding energies of H3K9ac and H3K27ac with wild-type and mutant YEATS domains (Table 1). We note that the MM/GBSA method used here involves several approximations, including mean-field solvent approximation and exclusion of configurational entropy, and therefore care must be taken while interpreting the results, which are largely enthalpic in nature. The binding energies obtained from MM/GBSA analysis suggest that H3K9ac interacted more favorably than H3K27ac with wild-type and deletion mutant, which is consistent with previous experimental studies.[12] The energetic analysis also showed that H3K9ac can stably bind to the insertion mutant, in

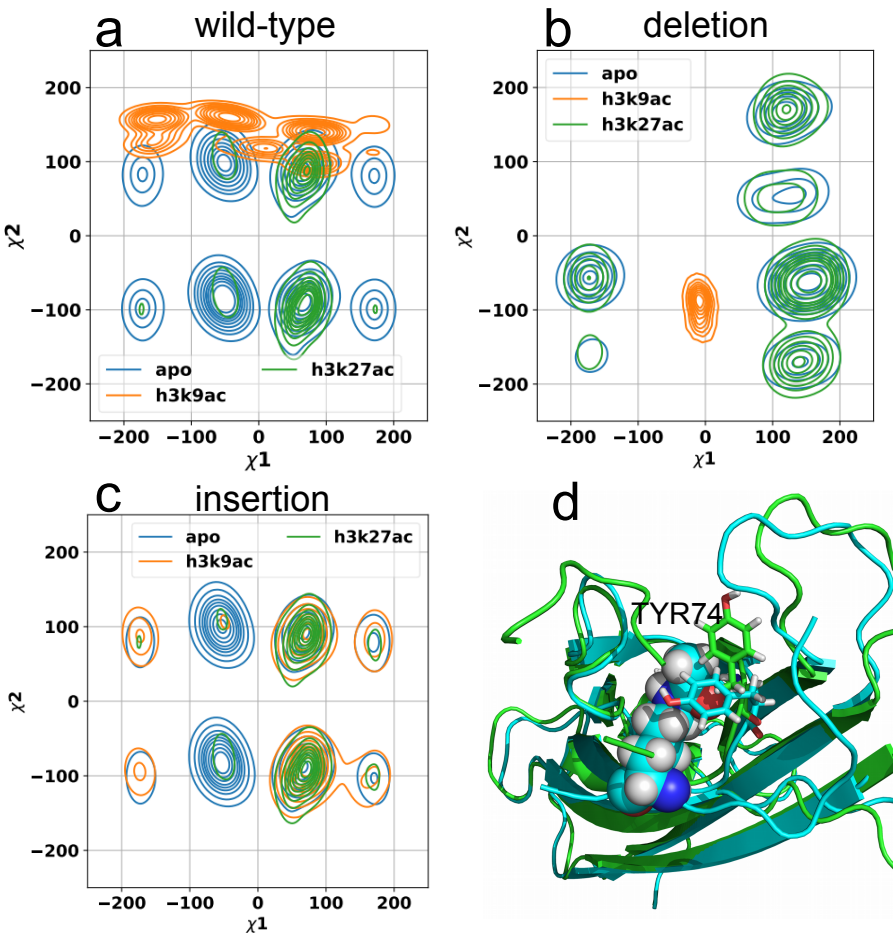

Figure 3: Distribution of side-chain torsional angles ($\chi1$, $\chi2$) of Tyr74 representing different conformational states of Tyr74 in apo, H3K27ac and H3K9ac bound wild-type, insertion and deletion mutant. a) wild-type YEATS, b)deletion-mutant, and c) insertion-mutant. d) Snapshot representing stacked (cyan) and unstacked (green) conformation of Tyr74 (shown in stick representation).

contrast to experimental studies which showed no detectable binding between H3K9ac and this mutant.[12,16]

H3K9ac and H3K27ac have similar residues in the consensus motif (−ARKacS−) but differ mainly in the N-terminal region. H3K9ac is more positively charged due to the presence of an additional arginine in the N-terminus. To understand the role of the amino-acid composition of the peptides on binding energies, we analyzed the per-residue contributions in the binding energies which revealed that the presence of Arg2 can provide modest energetic stabilization of ∼1.2 kcal/mol in H3K9ac as opposed to the negligible contribution of Leu2 in

Table 1: MM/GBSA derived binding energies of H3K9ac and H3K27ac with wild-type and mutant YEATS domains. All values are given in kcal/mol with standard error of the mean calculated across 10 simulations. Binding energies presented here should be interpreted comparatively, rather than absolute quantities.

Peptide	wild-type	Deletion	Insertion
H3K9ac	-43.94±1.05	-45.58±0.72	-49.38±1.01
H3K27ac	-40.27±0.28	-40.26±0.41	-41.94±0.45

H3K27ac at the same position (Figures 4a, and 4b). Further, additional residues present in H3K27ac (Ala6, Lys5) and H3K9ac (Thr6, Gln5) adjacent to the consensus motif provided higher energetic stabilization in the insertion mutant than the wild-type and deletion. This explains the higher number of contacts and binding energies of both peptides to the insertion mutant (Figure 4c).

In addition, we noted different levels of flexibility in H3K9ac and H3K27ac in the bound state. We examined the motion of peptides relative to the YEATS domain in all systems, which is an indirect measurement to assess the conformational freedom of histone peptides upon binding. The N-terminal region of the H3K9ac and H3K27ac in the YEATS domain fluctuated with RMSF values in the range of 4-10 Å and 5-12 Å, respectively, displaying higher flexibility of H3K27ac. The C-terminal region of the H3K27ac fluctuated in the range of 3-5 Å, whereas H3K9ac fluctuated in the range of 2-3 Å(Figure S4).We also observed a marginal increase in the flexibility of C-terminal region of H3K27ac when bound to deletion mutant. These RMSF differences highlight the role of peptide chemistry and mutations in modulating the peptide dynamics in the YEATS complex.

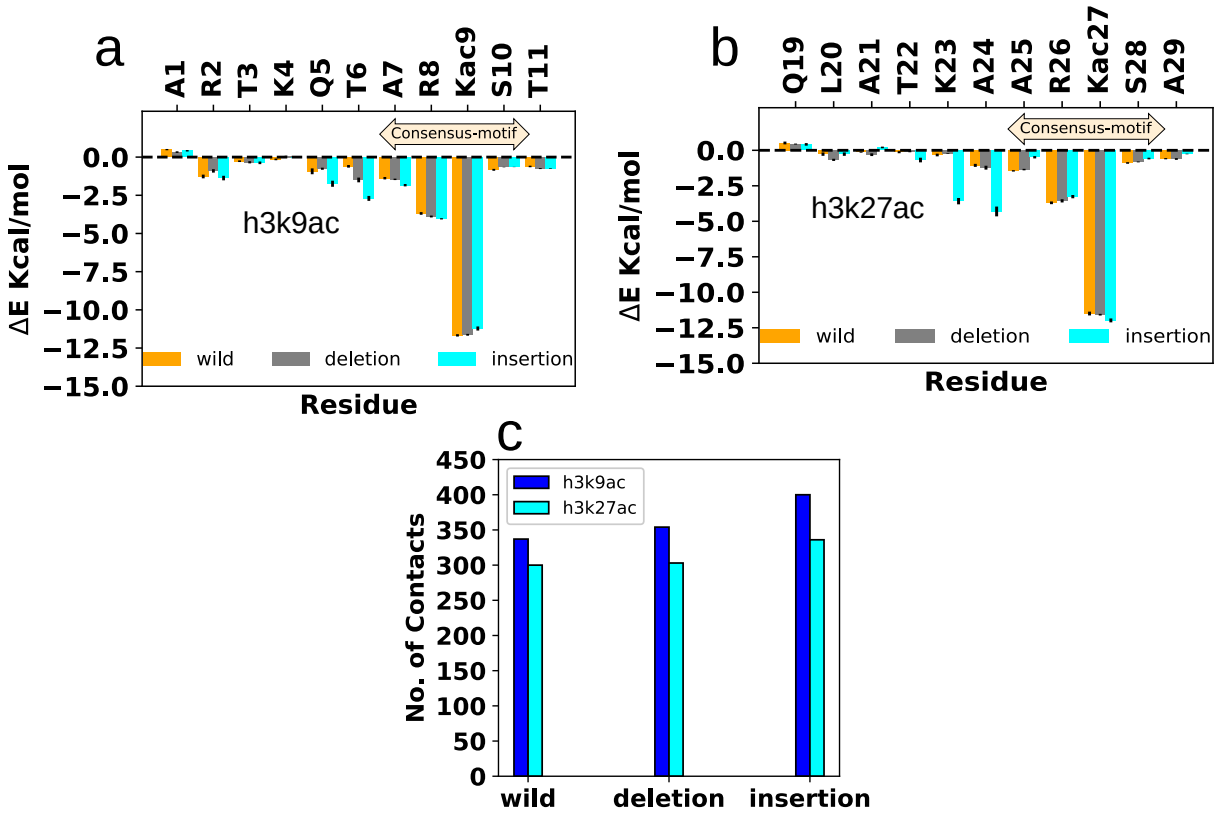

Figure 4: MM/GBSA derived per-residue binding energies of peptide mimics in complex with wild-type and mutant YEATS domains a) H3K9ac, b) H3K27ac, and c) Contacts formed between histone peptides (H3K9ac and H3K27) heavy atoms and wild-type and mutant YEATS domain within a distance cut-off of 4.5 Å.

Peptide binding results in conformational restrictions

Principal component analysis (PCA) was conducted on the conformational region conserved across the YEATS domains to dissect the large-scale motions and characterize the essential dynamics of YEATS apo and holo states (Figure 5). Conformational ensembles from all simulations were combined in the PCA to ensures that identical eigenbases were used in comparisons across the YEATS systems. In this analysis, Principal Component 1 (PC1) and PC2 represent the rocking motion and twisting motion of the YEATS domain, respectively (Figure 5d, 5e). Projections of the unbound wild-type simulations on PC1 ranged between -1.0 nm and 1.0 nm, whereas PC1 projections from both mutants show a marginal increase in a positive direction (PC1 -1.0 nm to 1.5 nm), indicating the absence of significant con-

formational changes upon mutations. Peptide binding only slightly restricts the motion of the YEATS domain in the PC1 direction, confirming the similar amount of conformational dynamics relative to the apo-states, which is in agreement with the RMSF analysis. In the PC2 space we noted a remarkable shift in the H3K9ac bound deletion mutant where the majority of the conformations lie in the positive direction as compared to the other system. This shift suggests a change in the global motion of the deletion mutant/H3K9ac system along with an inward twisting motion as opposed to outward twisting motion present in other systems.

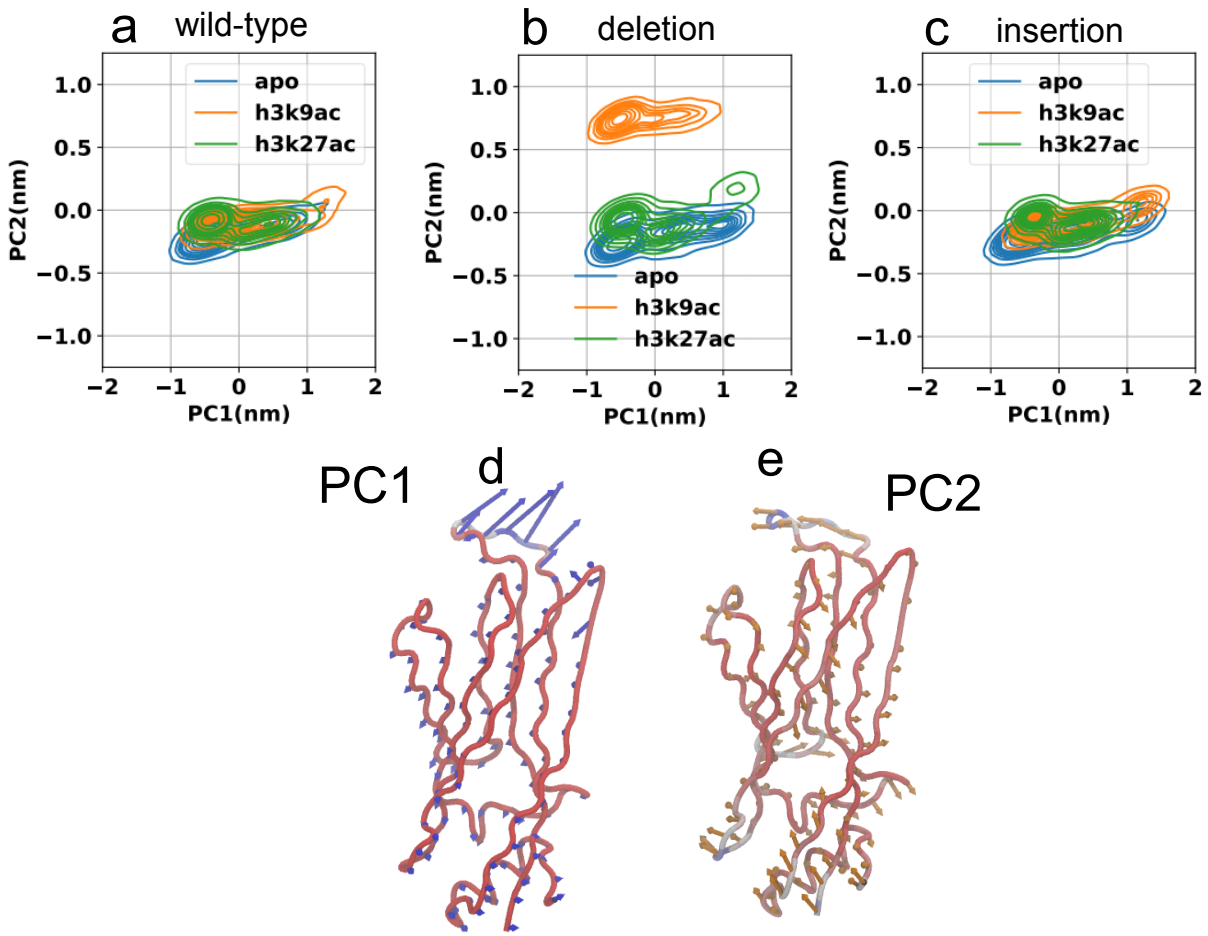

Figure 5: Conformational restrictions in peptide bound states of YEATS domain displayed by the motion along PC1 and PC2. PC2 suggested a inward twisting motion in deletion mutant (b) upon the binding of H3K9ac. PC1 suggested only minor variation in the conformational dynamics of wild-type (a) and insertion mutant(c). Procupine plots showing the direction of the motion of Cα atoms in YEATS domain along PC1 and PC2 representing the rocking and twisting motions.

Mutations weaken the inter-residue correlation

To understand how mutations can modulate inter-residue communication, the mutual information of Cα atoms in peptide bound and unbound states of each system was computed (Figure 6). The thick lines colored in red perpendicular to the diagonal in the plots represent the short-range correlation between anti-parallel β-strands, while dense irregular blocks show correlations between β-sheets (Figure 6a-c). β-strands and inter-β-sheet regions in wild-type YEATS exhibit a strong correlation, which was reduced upon insertion and deletion of residues in loop-8 (Figure 6a-c). This reduced correlation is likely due to the increased flexibility of mutants relative to the wild-type (Figures 2 & S1), which were highest for the insertion mutant and resulted in decoupling of short-range correlations and alterations of long-range correlations.

To understand how peptide-binding modulate the short and long-range correlations, the difference between apo and peptide (H3K9ac and H3K27ac)-bound correlations was calculated (Figure 6d-f & S5). The wild-type YEATS-bound H3K9ac system showed a reduced coupling in both local and distant regions, whereas the dynamics become more correlated between distant regions in the deletion mutant, although there was a net reduction in short-range correlations. In contrast, the dynamic coupling was largely unaffected by the binding of H3K9ac to the insertion mutant.

H3K27ac binding had a less drastic effect on the inter-residue dynamic correlation in the wild-type and deletion mutant relative to their apo-states. In the deletion mutant there were no regions in which there was an increase in correlations upon H3K27ac binding, and there were several regions in which there was a decrease in correlations (see Figure S5). Moreover, the coupling between structural regions in the insertion mutant remained unaffected upon H3K27ac binding (Figure S5). The changes observed in the correlation pattern across the apo and peptide bound states of YEATS were highly dependent on the peptide sequence and the type of mutation present in the YEATS.

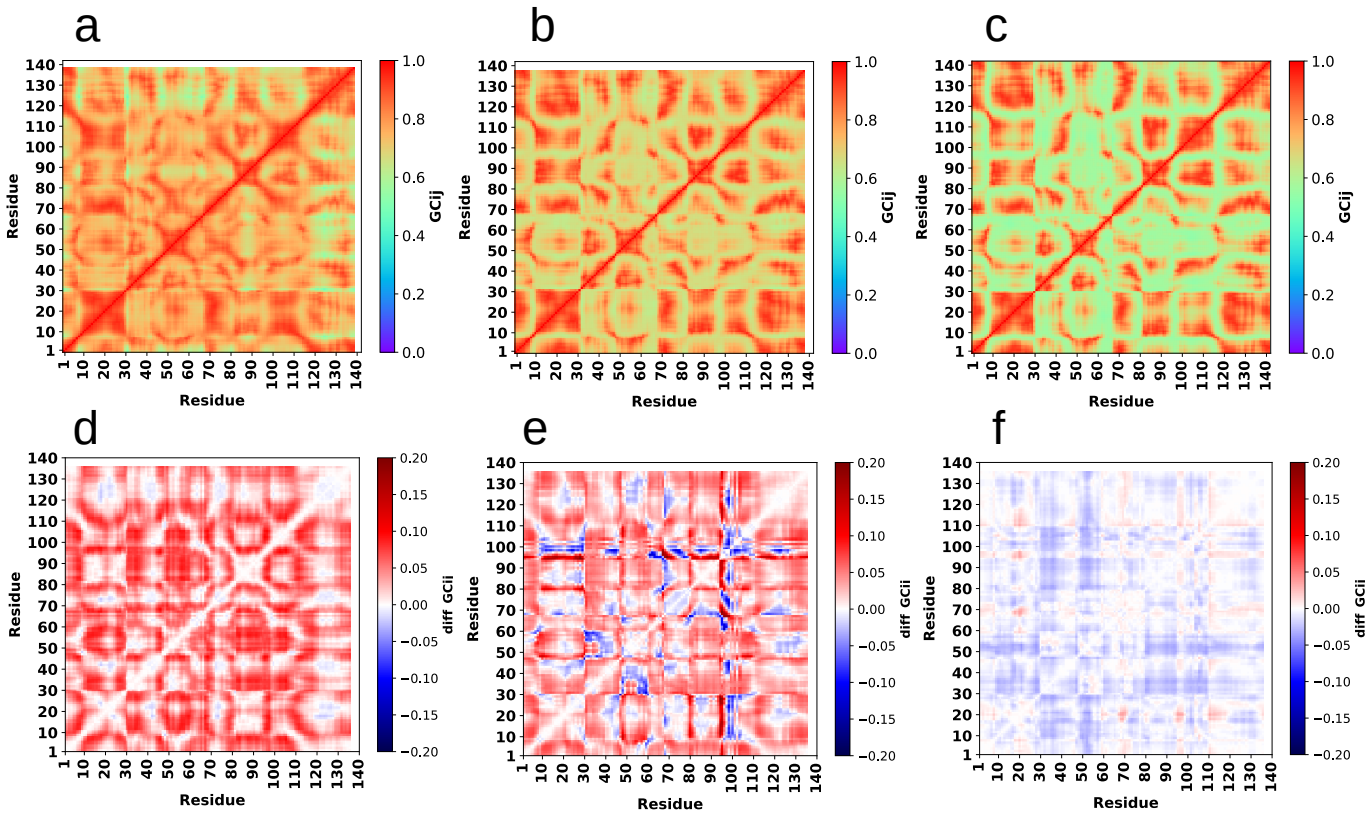

Figure 6: Decreased correlation in apo states of deletion (b) and insertion mutant (c) relative to the wild-type (a). Difference of mutual information of H3K9ac bound states to the apo-states of YEATS (d) wild-type (e) deletion mutant and (f) insertion mutant.

Discussion

Mutations in ENL YEATS are linked to altered binding abilities to histone acetylation marks.[12,16] To date, the structural basis of these altered binding mechanisms, and how they influence the binding thermodynamics and specificities of these systems, remains poorly understood, as structural structure have yet to reveal the binding mechanism of the full-length peptides used in ITC experiments. This has significantly affected our understanding of how the shift in binding preferences of mutant YEATS forms for acetylation sites on histone tails observed in binding assays translates to its altered chromatin occupancy. Here, we have used MD simulations to discern the differences in the binding of H3K9ac and H3K27ac with wild-type and mutant YEATS domain. Simulations of apo and bound states display only minor variations in the β-sheet core, with L8 having the greatest length and dynamics in

the insertion mutant, followed by the wild-type, and then the deletion mutant. Interestingly, peptide binding does not affect the dynamics of L8, which showed a similar level of dynamics relative to their apo forms (Figure 2). However, we observed system-specific changes in the structural elements in YEATS. For example,the dynamics of L1 are reduced in the deletion mutant upon H3K9ac binding, which is further confirmed by PCA analyses which showed a drastic shift in the PC2 subspace of the H3K9ac bound deletion mutant (Figure 5). We also noted a significant changes in the deletion mutant's internal motions upon H3K9ac binding relative to its apo and H3K27ac bound states, suggesting that H3K9ac induces unique conformational arrangements in the deletion mutant (Figure 3).

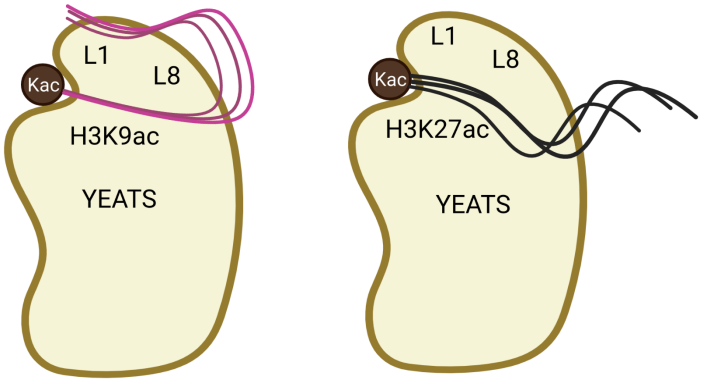

Figure 7: Representation showing the overall differences in the binding of H3K9ac and H3K27ac to YEATS domain. Non-consensus motif of H3K9ac binds with L1 loop, whereas H3K27ac binds more dynamically to YEATS domain.

Based on modeled and X-ray derived structures, it has been speculated that the H3K27ac and H3K9ac peptide mimics cannot bind to the YEATS insertion mutant because of steric hindrance from L8.[16] In contrast, ITC experiments showed that the deletion mutant had different thermodynamic signatures for each peptide's binding relative to the wild-type.[12,15,16] Here, we noted differences in the structural and dynamical properties of the H3K27ac and H3K9ac peptides which help to explain these disparate effects. H3K9ac binds to wild-type and mutant proteins with stronger affinities than H3K27ac due to the presence of an arginine in the non-consensus region of the peptide (Figure 4) that enabled H3K9ac to interact with

the L1 loop (Figure 7, and S6). Qualitative interpretations from our MM/GBSA analysis are in line with ITC experiments by Perlman *et al.,* which show a three times greater binding affinity of H3K9ac over H3K27ac.[12] Moreover, Ni *et al.,* reported a favorable enthalpic contribution in H3K9ac binding with the deletion mutant as opposed to an entropically favorable binding of H3K9ac with the wild-type.[16] Contact analysis reveals that the H3K9ac peptide forms a higher number of contacts with the deletion mutant, resulting in more conformational restrictions than in the wild-type and suggesting an increased role of enthalpic contributions in binding (Figure 4c).

The energetic analysis of our simulations showed that the insertion mutant can bind to both H3K9ac and H3K27ac. Our prediction for H3K27ac binding to the insertion mutant is in consensus with the reported binding of the insertion mutant to a H3K27ac peptide mimic.[15] Furthermore, per-residue energetic calculations showed that the binding to the YEATS domain is majorly dominated by the −ARKacS− consensus motif present in both peptides. The arginine in the consensus motif of the peptide tethers it along the β-sheet by forming a salt-bridge with YEATS' Asp99 residue (Figure S7). This study reveals that the peptide sequence beyond −ARKacS− can also make a difference in the overall affinity of the peptide mimics with the YEATS domain despite its more dynamic binding (Figure S4). The dynamic binding mode may preclude the binding information of full-length peptides from X-ray based studies.[11] However, MM/GBSA analysis alone cannot reproduce the absence of binding of H3K9ac to the insertion mutant. If unbinding of H3K9ac were to occur, it likely requires several microseconds. We did note in one of 10 simulations a major destabilization of the H3K9ac peptide to the insertion mutant, although extension of that simulation resulted in the peptide rebinding into its original position.

It has been shown in a recent study that despite weaker binding affinity of the insertion mutant to H3K27ac, it can robustly bind to acetylated chromatin.[15] Recent studies on the ENL analogue AF9 containing YEATS domain that were performed in the context of a full nucleosome showed an additive role of acetylation and the engagement of two YEATS

molecules on the same histone tails.[46] This study also highlighted the engagement of YEATS domain with a small region of the histone tail when YEATS interacts with an acetylated nucleosome core particle.[46–48] It has previously been shown that the exposure of histone tails in the context of the full nucleosome is limited by histone tail/DNA interactions.[17,49,50] Our MD studies also support that only partial exposure of the histone tails may be required for their engagement with a YEATS domain, and that the increased L8 length of YEATS may not prevent the engagement of the insertion mutant to H3K9ac mark on histone tails. Further computational and experimental studies are needed in the context of full length nucleosomes to understand how mutations linked with diseases modulates the preference of YEATS domains for lysine acetylation marks on histone tails.

Data and Software availability

Details for all the data generated in this molecular dynamics study are available in the Methods section including PDB codes of input files[11,16] and simulation protocol. The simulation[22,23] and visualization software[34,35] and the force field[24] used to perform this study are openly available. MD trajectories generated in this study are available on Zenodo at: `https://doi.org/10.5281/zenodo.6870475`.

Acknowledgement

The authors thank members of the Wereszczynski Group for valuable discussions concerning this work. This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health Grant R35GM119647 and the National Science Foundation grant CAREER-1552743. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

(1) Andrews, F. H.; Strahl, B. D.; Kutateladze, T. G. Insights into newly discovered marks and readers of epigenetic information. *Nature chemical biology* **2016**, *12*, 662–668.

(2) Erb, M. A.; Scott, T. G.; Li, B. E.; Xie, H.; Paulk, J.; Seo, H.-S.; Souza, A.; Roberts, J. M.; Dastjerdi, S.; Buckley, D. L., et al. Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **2017**, *543*, 270–274.

(3) Schulze, J. M.; Wang, A. Y.; Kobor, M. S. YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochemistry and Cell Biology* **2009**, *87*, 65–75.

(4) Hetzner, K.; Garcia-Cuellar, M.-P.; Büttner, C.; Slany, R. K. The interaction of ENL with PAF1 mitigates polycomb silencing and facilitates murine leukemogenesis. *Blood, The Journal of the American Society of Hematology* **2018**, *131*, 662–673.

(5) Garnar-Wortzel, L.; Bishop, T. R.; Kitamura, S.; Milosevich, N.; Asiaban, J. N.; Zhang, X.; Zheng, Q.; Chen, E.; Ramos, A. R.; Ackerman, C. J., et al. Chemical inhibition of ENL/AF9 YEATS domains in acute leukemia. *ACS central science* **2021**, *7*, 815–830.

(6) Li, X.; Liu, S.; Li, X.; Li, X. D. YEATS Domains as Novel Epigenetic Readers: Structures, Functions, and Inhibitor Development. *ACS Chemical Biology* **2022**,

(7) Heidenreich, D.; Moustakim, M.; Schmidt, J.; Merk, D.; Brennan, P. E.; Fedorov, O.; Chaikuad, A.; Knapp, S. Structure-based approach toward identification of inhibitory fragments for eleven-nineteen-leukemia protein (ENL). *Journal of Medicinal Chemistry* **2018**, *61*, 10929–10934.

(8) Moustakim, M.; Christott, T.; Monteiro, O. P.; Bennett, J.; Giroud, C.; Ward, J.; Rogers, C. M.; Smith, P.; Panagakou, I.; Díaz-Sáez, L., et al. Discovery of an MLLT1/3

YEATS domain chemical probe. *Angewandte Chemie International Edition* **2018**, *57*, 16302–16307.

(9) Zhao, D.; Li, Y.; Xiong, X.; Chen, Z.; Li, H. YEATS domain—A histone acylation reader in health and disease. *Journal of Molecular Biology* **2017**, *429*, 1994–2002.

(10) Andrews, F. H.; Shanle, E. K.; Strahl, B. D.; Kutateladze, T. G. The essential role of acetyllysine binding by the YEATS domain in transcriptional regulation. *Transcription* **2016**, *7*, 14–20.

(11) Wan, L.; Wen, H.; Li, Y.; Lyu, J.; Xi, Y.; Hoshii, T.; Joseph, J. K.; Wang, X.; Loh, Y.-H. E.; Erb, M. A., et al. ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* **2017**, *543*, 265–269.

(12) Perlman, E. J.; Gadd, S.; Arold, S. T.; Radhakrishnan, A.; Gerhard, D. S.; Jennings, L.; Huff, V.; Guidry Auvil, J. M.; Davidsen, T. M.; Dome, J. S., et al. MLLT1 YEATS domain mutations in clinically distinctive Favourable Histology Wilms tumours. *Nature communications* **2015**, *6*, 1–10.

(13) Gadd, S.; Huff, V.; Walz, A. L.; Ooms, A. H.; Armstrong, A. E.; Gerhard, D. S.; Smith, M. A.; Auvil, J. M. G.; Meerzaman, D.; Chen, Q.-R., et al. A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nature genetics* **2017**, *49*, 1487–1494.

(14) Doty, R. T.; Vanasse, G. J.; Disteche, C. M.; Willerford, D. M. The leukemia-associated gene Mllt1/ENL: characterization of a murine homolog and demonstration of an essential role in embryonic development. *Blood Cells, Molecules, and Diseases* **2002**, *28*, 407–417.

(15) Wan, L.; Chong, S.; Xuan, F.; Liang, A.; Cui, X.; Gates, L.; Carroll, T. S.; Li, Y.; Feng, L.; Chen, G., et al. Impaired cell fate through gain-of-function mutations in a chromatin reader. *Nature* **2020**, *577*, 121–126.

(16) Ni, X.; Londregan, A. T.; Owen, D. R.; Knapp, S.; Chaikuad, A. Structure and inhibitor binding characterization of oncogenic MLLT1 mutants. *ACS Chemical Biology* **2021**, *16*, 571–578.

(17) Morrison, E. A.; Bowerman, S.; Sylvers, K. L.; Wereszczynski, J.; Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **2018**, *7*, e31481.

(18) Marunde, M. R.; Fuchs, H. A.; Burg, J. M.; Popova, I. K.; Vaidya, A.; Hall, N. W.; Meiners, M. J.; Watson, R.; Howard, S. A.; Novitzky, K., et al. Nucleosome conformation dictates the histone code. *BioRxiv* **2022**,

(19) Ghoneim, M.; Fuchs, H. A.; Musselman, C. A. Histone tail conformations: a fuzzy affair with DNA. *Trends in biochemical sciences* **2021**, *46*, 564–578.

(20) Shen, M.-y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein science* **2006**, *15*, 2507–2524.

(21) Li, Y.; Wen, H.; Xi, Y.; Tanaka, K.; Wang, H.; Peng, D.; Ren, Y.; Jin, Q.; Dent, S. Y.; Li, W., et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **2014**, *159*, 558–571.

(22) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of chemical theory and computation* **2012**, *8*, 1542–1555.

(23) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Computer Physics Communications* **2013**, *184*, 374–380.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmer-

ling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.

(25) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: a different approach. *The journal of physical chemistry letters* **2014**, *5*, 3863–3871.

(26) Papamokos, G. V.; Tziatzos, G.; Papageorgiou, D. G.; Georgatos, S. D.; Politou, A. S.; Kaxiras, E. Structural role of RKS motifs in chromatin interactions: a molecular dynamics study of HP1 bound to a variably modified histone tail. *Biophysical journal* **2012**, *102*, 1926–1933.

(27) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of chemical theory and computation* **2015**, *11*, 1864–1874.

(28) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* **1977**, *23*, 327–341.

(29) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical physics letters* **2004**, *384*, 288–294.

(30) Lange, O. F.; Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, and Bioinformatics* **2006**, *62*, 1053–1061.

(31) Dai, J.; Yuan, J. Large-scale efficient Langevin dynamics, and why it works. *EPL (Europhysics Letters)* **2009**, *88*, 20001.

(32) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *Journal of Chemical Theory and Computation* **2018**, *14*, 6127–6138.

(33) Braun, E.; Gilmer, J.; Mayes, H.; Mobley, D.; Monroe, J.; Prasad, S.; Zuckerman, D. Best Practices for Foundations in Molecular Simulations [Article v1. 0]. Living J Comp Mol Sci. 2018; 1 (1): 5957.

(34) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14*, 33–38.

(35) DeLano, W. L., et al. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **2002**, *40*, 82–92.

(36) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **2013**, *9*, 3084–3095.

(37) McClendon, C. L.; Hua, L.; Barreiro, G.; Jacobson, M. P. Comparing conformational ensembles using the Kullback–Leibler divergence expansion. *Journal of chemical theory and computation* **2012**, *8*, 2115–2126.

(38) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying correlations between allosteric sites in thermodynamic ensembles. *Journal of chemical theory and computation* **2009**, *5*, 2486–2502.

(39) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry* **2011**, *32*, 2319–2327.

(40) Miller III, B. R.; McGee Jr, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA. py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation* **2012**, *8*, 3314–3321.

(41) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the performance of MM/PBSA

and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *The journal of physical chemistry B* **2013**, *117*, 8408–8421.

(42) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics* **2008**, *129*, 124105.

(43) David, C. C.; Jacobs, D. J. *Protein dynamics*; Springer, 2014; pp 193–226.

(44) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005**, *26*, 1701–1718.

(45) Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Physical review E* **2004**, *69*, 066138.

(46) Kikuchi, M.; Morita, S.; Goto, M.; Wakamori, M.; Katsura, K.; Hanada, K.; Shirouzu, M.; Umehara, T. Elucidation of binding preferences of YEATS domains to site-specific acetylated nucleosome core particles. *Journal of Biological Chemistry* **2022**, 102164.

(47) Klein, B. J.; Vann, K. R.; Andrews, F. H.; Wang, W. W.; Zhang, J.; Zhang, Y.; Beloglazkina, A. A.; Mi, W.; Li, Y.; Li, H., et al. Structural insights into the π-π-π stacking mechanism and DNA-binding activity of the YEATS domain. *Nature communications* **2018**, *9*, 1–11.

(48) Weaver, T. M.; Morrison, E. A.; Musselman, C. A. Reading more than histones: the prevalence of nucleic acid binding among reader domains. *Molecules* **2018**, *23*, 2614.

(49) Vann, K. R.; Klein, B. J.; Kutateladze, T. G. Mechanistic similarities in recognition of histone tails and DNA by epigenetic readers. *Current Opinion in Structural Biology* **2021**, *71*, 1–6.

(50) Peng, Y.; Li, S.; Onufriev, A.; Landsman, D.; Panchenko, A. R. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails. *Nature communications* **2021**, *12*, 1–12.