# Trimethyllysine reader proteins exhibit widespread charge-agnostic binding via different mechanisms to cationic and neutral ligands

Christopher R. Travis, Kelsey M. Kean, Katherine I. Albanese, Hanne C. Henriksen, Joseph W. Treacy, Elaine Y. Chao, K. N. Houk, Marcey L. Waters*

**ABSTRACT:** In the last 40 years, cation–π interactions have become part of the lexicon of noncovalent forces that drive protein binding. Indeed, tetraalkylammoniums are universally bound by aromatic cages in proteins, suggesting that cation–π interactions are a privileged mechanism for binding these ligands. A prominent example is the recognition of histone trimethyllysine (Kme3) by the conserved aromatic cage of reader proteins, dictating gene expression. However, two proteins have recently been suggested as possible exceptions to conventional understanding of tetraalkylammonium recognition. To broadly interrogate the role of cation–π interactions in protein binding interactions, we report the first large-scale comparative evaluation of reader proteins for a neutral Kme3 isostere, experimental and computational mechanistic studies, and structural analysis. We find unexpected widespread binding of readers to a neutral isostere, with no single factor dictating charge selectivity, demonstrating the challenge to predict such interactions. Further, readers that bind both cationic and neutral ligands display an unprecedented change in mechanism: binding Kme3 via cation–π interactions and the neutral isostere through the hydrophobic effect in the same aromatic cage. This discovery challenges traditional understanding of molecular recognition of tetraalkylammoniums by aromatic cages in myriad protein-ligand interactions and establishes a new framework for selective inhibitor design by exploiting differences in charge-dependence.

## INTRODUCTION

Methyllysine (Kme) reader proteins bind to this key post-translational modification (PTM) on histone tails in a sequence-selective manner, resulting in recruitment of proteins to the nucleosome that dictate gene expression and elicit downstream effects.[1-7] Given their significant roles in epigenetic pathways, dysregulation of histone Kme readers is associated with numerous diseases, and many of these proteins have emerged as therapeutic targets.[8-10] Approximately 200 human histone Kme readers have been characterized to date,[1,2,11,12] many of which bind the same PTM at the identical histone tail residue, yet elicit distinct biological outcomes.[13]

Several reports describe medicinal chemistry efforts targeting Kme readers.[14-19] However, selectivity remains a major obstacle, as all known Kme3 readers bind the quaternary ammonium group within a conserved aromatic cage as well as through interactions with the surrounding sequence (Figure 1a).[1,2] To date, no therapeutics targeting histone Kme reader proteins have been approved for treatment.[20-22] Given the well-documented challenges faced in developing selective inhibitors for this class of proteins,[22-26] a thorough understanding of the binding capabilities of the aromatic cages of Kme3 readers is warranted to provide insight into novel approaches to achieve selective inhibition.

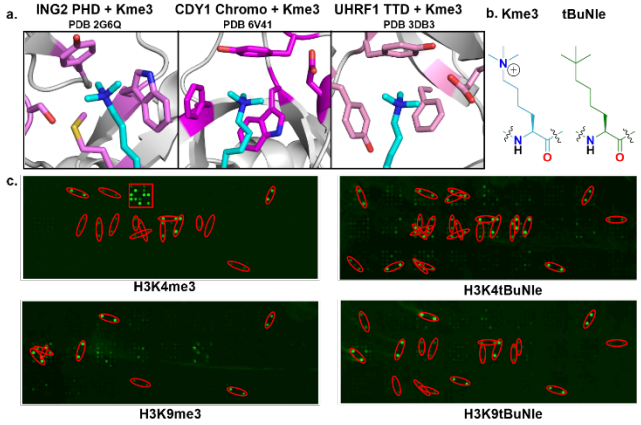



**Figure 1.** (a) Aromatic cages of three readers from different families evaluated in this study. (b) Structures of Kme3 and tBuNle. (c) Histone methyllysine reader protein microarray shows many reader proteins from various families are capable of binding the tBuNle mark at H3K4 and H3K9 (significant interactions indicated in red.

For nearly 40 years, mechanistic studies of the molecular recognition of quaternary ammonium ions, including Kme3, have repeatedly demonstrated that aromatic cages in proteins as well as synthetic systems preferentially bind cationic ligands over neutral ligands, emphasizing the critical importance of cation–π interactions to these binding events.[27-35] In accordance with these studies, maintaining cation–π interactions has been a standard approach to inhibitor development for Kme readers,[22,26,36] as well as other trimethylammonium-binding proteins.[28,29] However, recently a few isolated examples have been reported that contradict the preference of aromatic cages for cationic ligands.[37-39] Moreover, conflicting mechanistic explanations have been proposed, such as binding being driv-

en by displacement of high energy water molecules in the aromatic cage regardless of the nature of the ligand.[37] Because of the biological and medicinal relevance of these interactions as well as the fundamental significance of cation–π interactions in mediating recognition of tetraalkylammoniums in biological systems, we undertook a comprehensive screen of Kme reader protein selectivity for Kme3 vs its neutral isostere tert-butyl norleucine (tBuNle) coupled with mechanistic and structural studies to determine the driving force for aromatic cages binding cationic vs neutral ligands. We find that 10 out of 20 proteins identified from the screen bind to the uncharged tBuNle with affinity comparable to or tighter than to Kme3. Furthermore, we find that readers that bind both marks do so through an unprecedented switch in mechanism, binding Kme3 via cation–π interactions and tBuNle via the hydrophobic effect. Finally, we find that there is no one structural feature that dictates selectivity, meaning that structural analysis or *in silico* screening methods would not predict our findings. Together, these findings demonstrate a complex molecular recognition profile by aromatic cages that bind trimethylammoniums and explains apparently contradictory results in the literature.[27,37] Moreover, this work enables novel strategies to selectively target a subset of aromatic-cage containing proteins that are charge-agnostic in binding.

## RESULTS AND DISCUSSION

**Many methyllysine readers bind histone tails containing a neutral Kme3 analog.** To determine the scope of histone Kme readers that bind the neutral isostere of Kme3, we performed a qualitative screen of biotinylated H3 peptides, each containing Kme3 or tBuNle at the K4 or K9 position, against a reader protein microarray that contains approximately 200 different human proteins that are known or predicted binders of histone methylation (Figure 1c, Figures S1, S2)[16] Results indicate that many histone Kme readers bind to H3K4tBuNle and H3K9tBuNle, suggesting the ability of readers to bind a neutral ligand is more widespread than previously known.[37,38]

Twenty histone Kme readers were selected for quantitative analysis of binding to Kme3 and tBuNle histone peptides via isothermal titration calorimetry (ITC) (Figures S3, S4, S5). We found selectivities ranging from > 120:1 in favor of Kme3 to nearly 7:1 in favor of tBuNle among these proteins (Table 1, Figure S6). The degree of selectivity differs among families: all chromodomains investigated preferentially bind Kme3 whereas readers from the plant homeodomain (PHD) and Tudor domain (TTD) families exhibit a range of selectivities. Eight PHDs bind tBuNle with comparable or tighter affinity than Kme3 (DIDO1, RAG2, BPTF, and ING1/2/3/4/5), while three members of this family demonstrate greater than 2-fold preference for Kme3 (TAF3, JARID1A, and JARID1B). Among the TTDs, SGF29 preferentially binds tBuNle, while JMJD2A and SPIN1 have large preferences for Kme3. Further, we find that the UHRF1 TTD, a key oncological target,[40] binds H3K9tBuNle nearly 7-fold more tightly than H3K9me3, the largest degree of preference observed for tBuNle of any protein, as well as the only H3K9me3 reader to prefer tBuNle. These results are unexpected, as each of the readers investigated contains a highly conserved cage of two to four aromatic residues in which Kme3 binds, yet these proteins, even those in the same family, differ substantially in their affinities for tBuNle. Mechanistic studies and structural analyses were undertaken to assess factors which dictate this range of selectivities.

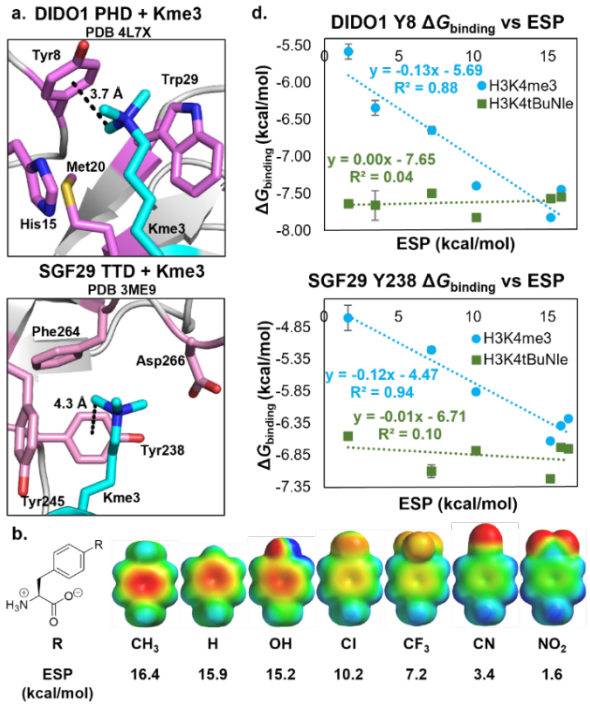
**Table 1.** Binding affinities and selectivities of histone readers binding to H3 peptides containing Kme3 and tBuNle determined by ITC at 25 °C, pH 7.4, 50 mM sodium phosphate, 150 mM (or 500 mM‡) NaCl, and 2 mM TCEP. *TAF3 PHD data was previously reported.[38] ^CBX1 chromodomain data with H3K9me3 was previously reported.[41]

| Protein | Ligand | $K_D$ (μM) | $K_D$(Kme3)/ $K_D$(tBuNle) | $\Delta G_{binding}$ (kcal/mol) |
|---|---|---|---|---|
| TAF3 PHD* | H3K4me3 | 1.1 ± 0.2 | 0.1 | -8.16 ± 0.02 |
| | H3K4tBuNle | 11 ± 1 | | -6.78 ± 0.04 |
| JARID1A PHD(3)‡ | H3K4me3 | 3.6 ± 0.1 | 0.46 | -7.43 ± 0.01 |
| | H3K4tBuNle | 7.9 ± 0.2 | | -6.96 ± 0.01 |
| JARID1B PHD(3)‡ | H3K4me3 | 3.8 ± 0.1 | 0.46 | -7.40 ± 0.01 |
| | H3K4tBuNle | 8.2 ± 0.3 | | -6.94 ± 0.01 |
| DIDO1 PHD | H3K4me3 | 1.9 ± 0.1 | 0.66 | -7.82 ± 0.01 |
| | H3K4tBuNle | 2.9 ± 0.2 | | -7.56 ± 0.04 |
| RAG2 PHD | H3K4me3 | 42 ± 6 | 0.74 | -5.98 ± 0.08 |
| | H3K4tBuNle | 57 ± 6 | | -5.79 ± 0.07 |
| BPTF PHD | H3K4me3 | 8 ± 1 | 1.2 | -7.0 ± 0.1 |
| | H3K4tBuNle | 6.6 ± 0.3 | | -7.07 ± 0.03 |
| ING3 PHD | H3K4me3 | 2.5 ± 0.1 | 1.7 | -7.64 ± 0.01 |
| | H3K4tBuNle | 1.5 ± 0.2 | | -7.96 ± 0.06 |
| ING2 PHD | H3K4me3 | 8.5 ± 0.5 | 2.2 | -6.92 ± 0.04 |
| | H3K4tBuNle | 3.8 ± 0.1 | | -7.40 ± 0.02 |
| ING1 PHD | H3K4me3 | 6.8 ± 0.7 | 2.8 | -7.05 ± 0.06 |
| | H3K4tBuNle | 2.4 ± 0.1 | | -7.67 ± 0.02 |
| ING4 PHD | H3K4me3 | 10.2 ± 0.6 | 3 | -6.81 ± 0.04 |
| | H3K4tBuNle | 3.4 ± 0.2 | | -7.46 ± 0.03 |
| ING5 PHD | H3K4me3 | 17.3 ± 0.7 | 3.3 | -6.50 ± 0.01 |
| | H3K4tBuNle | 5.2 ± 0.1 | | -7.21 ± 0.01 |
| CBX1 Chromo^ | H3K9me3 | 0.68 ± 0.08 | 0.017 | -8.42 ± 0.03 |
| | H3K9tBuNle | 40 ± 10 | | -6.07 ± 0.09 |
| CHD1 Chromo | H3K4me3 | 18.8 ± 0.7 | 0.13 | -6.45 ± 0.02 |
| | H3K4tBuNle | 150 ± 20 | | -5.2 ± 0.1 |
| CDY1 Chromo | H3K9me3 | 1.75 ± 0.08 | 0.13 | -7.85 ± 0.02 |
| | H3K9tBuNle | 14 ± 1 | | -6.63 ± 0.04 |
| CBX5 Chromo | H3K9me3 | 15 ± 4 | 0.15 | -6.6 ± 0.2 |
| | H3K9tBuNle | ~100 | | >-6 |
| MPP8 Chromo‡ | H3K9me3 | 2.5 ± 0.1 | 0.16 | -7.65 ± 0.02 |
| | H3K9tBuNle | 15.3 ± 0.6 | | -6.57 ± 0.03 |
| SPIN1 TTD‡ | H3K4me3 | 0.67 ± 0.04 | 0.0079 | -8.42 ± 0.01 |
| | H3K4tBuNle | 85 ± 10 | | -5.56 ± 0.07 |
| JMJD2A TTD‡ | H3K4me3 | 19.5 ± 0.4 | 0.039 | -6.42 ± 0.01 |
| | H3K4tBuNle | >500 | | >-5 |
| SGF29 TTD | H3K4me3 | 13.8 ± 0.6 | 2.7 | -6.63 ± 0.03 |
| | H3K4tBuNle | 5.1 ± 0.3 | | -7.22 ± 0.04 |
| UHRF1 TTD‡ | H3K9me3 | 44 ± 3 | 6.8 | -5.94 ± 0.03 |
| | H3K9tBuNle | 6.5 ± 0.8 | | -7.1 ± 0.1 |

**Kme3 and tBuNle bind by different mechanisms.** The cation–π interaction has been shown to be the key driving force for binding Kme3 in several readers.[27,37,41,42] Thus, for

proteins which bind tBuNle with equal or tighter affinity, the question arises as to the driving force of binding these two differently charged ligands. ITC data indicates that two proteins, the DIDO1 and ING3 PHDs, display enthalpy-entropy compensation (EEC), with similar free energies of binding to Kme3 and tBuNle, but distinct differences in enthalpy and entropy of binding (Figure S6). For these proteins, binding of Kme3 is more enthalpically favorable while binding of tBuNle has a significantly reduced enthalpic driving force and a favorable entropy of binding, consistent with a hydrophobic driving force. Other proteins, including the BPTF PHD, also bind tBuNle with comparable affinities, but do not demonstrate significant EEC (Figure S6). While the observed EEC for some proteins suggests possible differences in mechanism, these data alone are not sufficient to draw a definitive conclusion, as EEC is not a consistently reliable measure of the hydrophobic driving force.[43] Thus, we pursued a more rigorous method to evaluate the binding mechanisms of readers to Kme3 and tBuNle.

Electrostatic tunability of aromatic residues is an established tool to evaluate cation–π interactions.[44-48] Previously, we developed a genetic code expansion (GCE) methodology incorporating of *para*-substituted Phe derivatives[41] to systematically tune the electrostatic potentials (ESP) of aromatic residues within cages of readers, which provides information on the electrostatic contribution to binding at the individual residue level (Figure 2). This approach has elucidated the driving forces of various reader proteins binding to different PTMs,[49,50] including cation–π interactions with Kme3.[41,42] Using this GCE method, we investigated the DIDO1 PHD and the SGF29 TTD (Figure 2). We prepared a series of DIDO1 PHD Tyr8 mutants and SGF29 TTD Tyr238 mutants via GCE,[51] and measured binding to H3K4me3 and H3K4tBuNle peptides via ITC (Figure 2, Figure S5).




b.



| R | $CH_3$ | H | OH | Cl | $CF_3$ | CN | $NO_2$ |
|---|---|---|---|---|---|---|---|
| ESP (kcal/mol) | 16.4 | 15.9 | 15.2 | 10.2 | 7.2 | 3.4 | 1.6 |

c.

| Protein | H3K4me3 | | H3K4tBuNle | |
|---|---|---|---|---|
| | $K_D$ (μM) | $\Delta G_{binding}$ (kcal/mol) | $K_D$ (μM) | $\Delta G_{binding}$ (kcal/mol) |
| DIDO1 WT | 1.9 ± 0.1 | -7.82 ± 0.01 | 2.9 ± 0.2 | -7.56 ± 0.04 |
| DIDO1 Y8F | 3.5 ± 0.2 | -7.45 ± 0.03 | 3.0 ± 0.2 | -7.54 ± 0.04 |
| DIDO1 Y8*p*ClF | 3.9 ± 0.2 | -7.39 ± 0.03 | 1.9 ± 0.2 | -7.82 ± 0.02 |
| DIDO1 Y8*p*$CF_3$F | 14 ± 1 | -6.64 ± 0.05 | 3.2 ± 0.2 | -7.49 ± 0.02 |
| DIDO1 Y8*p*CNF | 23 ± 3 | -6.3 ± 0.1 | 2.5 ± 0.6 | -7.7 ± 0.2 |
| DIDO1 Y8*p*$NO_2$F | 80 ± 20 | -5.6 ± 0.1 | 2.6 ± 0.2 | -7.63 ± 0.04 |
| SGF29 WT | 13.8 ± 0.6 | -6.63 ± 0.03 | 5.1 ± 0.3 | -7.22 ± 0.04 |
| SGF29 Y238F | 21.0 ± 0.8 | -6.39 ± 0.02 | 11.6 ± 0.5 | -6.73 ± 0.03 |
| SGF29 Y238*p*$CH_3$F | 24.9 ± 0.6 | -6.28 ± 0.02 | 11.4 ± 0.4 | -6.75 ± 0.03 |
| SGF29 Y238*p*ClF | 51 ± 3 | -5.86 ± 0.04 | 11 ± 1 | -6.78 ± 0.06 |
| SGF29 Y238*p*$CF_3$F | 150 ± 10 | -5.20 ± 0.05 | 6 ± 1 | -7.1 ± 0.1 |
| SGF29 Y238*p*$NO_2$F | 380 ± 70 | -4.7 ± 0.2 | 16 ± 1 | -6.55 ± 0.04 |

**Figure 2.** (a) Aromatic cages of DIDO1 PHD and SGF29 TTD. (b) Structure of *para*-substituted Phe and ESP maps and values of analogs used in this study. ESP maps were calculated in Spartan using the DFT ωB97X-D/6-31G(d) level of theory with an energetic range from -100 to +100 kcal/mol. Blue indicates positive ESP, green is neutral, and red indicates negative ESP. ESP values were previously reported.[52] (c) Binding affinities of wild-type and *para*-Phe mutant DIDO1 PHD and SGF29 TTD to H3K4me3 and H3K4tBuNle determined by ITC. Each value reflects the average of three independent ITC experiments, with error as the larger of the standard deviation of the three runs or the largest error from an individual run. (d) LFER plots evaluating the correlation between $\Delta G_{binding}$ and calculated ESP values.

The DIDO1 PHD domain series of Tyr8 mutants exhibits significant differences in binding affinity to the Kme3 peptide, with electron-withdrawing groups resulting in weaker binding. Plotting the free energy of binding ($\Delta G_{binding}$) against the calculated ESP value[52] of each Tyr8 mutant results in a linear free energy relationship (LFER), indicating that binding is driven by electrostatically tunable cation–π interactions (Figure 2,

Figure S7). The slope of this interaction, which indicates the degree of electrostatic influence on binding, is comparable to previously studied readers binding Kme3.[41,42] In contrast, plotting $\Delta G_{binding}$ for tBuNle with the series of Tyr8 mutants against ESP results in a flat line, indicating that electrostatics have no influence on binding to this ligand (Figure 2). These results demonstrate that the DIDO1 PHD binds Kme3 via cation–π interactions as has been observed in other Kme3 readers but binds to tBuNle via the hydrophobic effect, which is independent of electrostatics, in agreement with observed EEC (Figure S6). Similarly, for the SGF29 TTD, reduction in the ESP in a series of Tyr238 mutants also correlates to decreased binding affinities to the Kme3 peptide, consistent with tunable cation–π interactions (Figure 2, Figure S7), whereas binding to tBuNle by the SGF29 TTD is not significantly affected by mutations at Tyr238. Together, GCE studies indicate that both proteins, which are from two different families, utilize different mechanisms to recognize Kme3 versus tBuNle within the same aromatic cage binding motif, even though the wild-type $\Delta G_{binding}$ to the two ligands is similar. To gain further insight, we performed energy decomposition analysis calculations on SGF29 TTD binding the two ligands at the M06-2X/6-311++G(d,p), SMD(Water) level of theory. These results demonstrate that binding to the two ligands differs primarily in the contribution of electrostatics and desolvation to the interaction, providing further support for different mechanisms of binding (Figure 3a, Figure S8). This finding, that the same aromatic cage binds cationic and neutral ligands with comparable affinities but by completely different mechanisms, is unprecedented.

Further analysis of the contribution of the cation–π interaction was accomplished through a double mutant cycle (Figure S9) using tBuNle as the Kme3 mutation, as it lacks a cation, and *p*$NO_2$Phe as the Tyr mutation, as it lacks a significant electrostatic contribution with an ESP value near zero (Figure 2b). This novel approach allows us to quantify cation–π interactions for a single aromatic residue. The contribution of the cation–π interaction to binding is estimated to be -2.3 ± 0.1 kcal/mol for Tyr8 in DIDO1 PHD and -1.3 ± 0.2 kcal/mol for Tyr238 in SGF29 TTD. This emphasizes that although these proteins bind with comparable or tighter affinity to the neutral ligand, cation–π interactions are still significant in binding Kme3. Moreover, these results emphasize that measuring binding of tBuNle to wild-type protein is not a sufficient mechanistic tool for probing the contribution of cation–π interactions.

**Neutral isostere binds in the same pocket as Kme3.** We solved a 1.3 Å resolution crystal structure of the RAG2 PHD bound to H3K4tBuNle (Figure 3b, PDB 8T4R) that exhibits a highly similar conformation to this protein binding Kme3 (PDB 2V89), emphasizing that tBuNle does not access a different binding site or contacts than Kme3 while binding with comparable affinity (Figure 3, Figure S10).[53] This is in agreement with previously determined structures with tBuNle that overlay with Kme3 (Figure S8).[37]

**Acidic residues in the binding pocket do not predict charge preference.** We evaluated whether the presence of an acidic residue near a reader's aromatic cage may drive selectivity for Kme3 through a favorable electrostatic interaction between the cationic ligand and the anionic acidic residue and/or unfavorable desolvation of the acidic residue upon binding of tBuNle.[54] Multiple Kme3 readers have an acidic residue within 5 Å of a methyl or methylene group on the ε-N of Kme3 and are selective for Kme3 (Figure S11). However, that is not sufficient to dictate a preference for Kme3. For example, the SGF29 and UHRF1 TTDs each contain an acidic residue in their aromatic cage 3.4 Å away from the nearest Kme3 methyl group but are not selective for Kme3 (Figures S8, S11, S12). Additionally, several proteins that are selective for Kme3 do not have an acidic residue in the cage, including the JARID1A/B PHDs and the CBX5 chromodomain (Figures S8, S11, S12). Thus, identification of an acidic residue in the aromatic cage does not determine selectivity or lack thereof for Kme3 over tBuNle.

**Calculated aromatic cage interactions do not predict charge preference.** Since each reader differs in the number and type of aromatic residues in its cage (Figures S8, S11, S12), and the strengths of cation–π interactions are affected by distance and angle,[30] we performed computational studies at the M06-2X/6-311++G(d,p), SMD(Water) level of theory. For each protein with an available structure, we calculated interaction energies ($E_{Int}$) between Kme3 or tBuNle and each amino acid in the binding pocket (Table 2, Figures S8, S12). Plotting the sum of the $E_{Int}$ for each interaction in the cage versus the experimental ΔΔG(Kme3/tBuNle) resulted in minimal correlation (Figures S11, S13), indicating that stronger structurally predicted cation–π interactions (i.e. larger summed $E_{Int}$) do not correlate with increased selectivity for Kme3.

**Interactions with the surrounding histone tail do not predict charge preference.** Given that Kme readers bind histones through cooperative interactions with both Kme3 and the surrounding sequence, we considered whether readers which have lower selectivity for histones containing Kme3 over unmodified lysine may also be less selective for Kme3 over tBuNle. Comparison of the ratio of the $K_D$ of binding to Kme3 versus the unmodified H3 tail (based on literature values, Figure S11) to the ratio of binding to Kme3 and tBuNle shows no correlation, suggesting that differences in binding to the sequence surrounding Kme3 do not dictate reader protein selectivity for Kme3 vs tBuNle (Figure S14). Furthermore, the five ING PHDs, which have nearly identical aromatic cages (Figure S15) but differ in residues that interface with the surrounding histone tail,[55] demonstrate similar selectivity profiles (Table 1), indicating that differences in interactions with the histone tail outside of the aromatic cage do not dictate preference for Kme3 vs tBuNle.

In sum, no single unifying feature dictates selectivity for Kme3 vs tBuNle. Instead, differences in many contributing interactions, which amount to a sum of small differences in $\Delta G_{binding}$, contribute to the range of selectivities observed for the 20 proteins that were investigated. This work elucidates that the aromatic cages of Kme3 readers are significantly more complex than previously thought and that structural analysis is not sufficient to predict selectivity for Kme3 vs tBuNle.

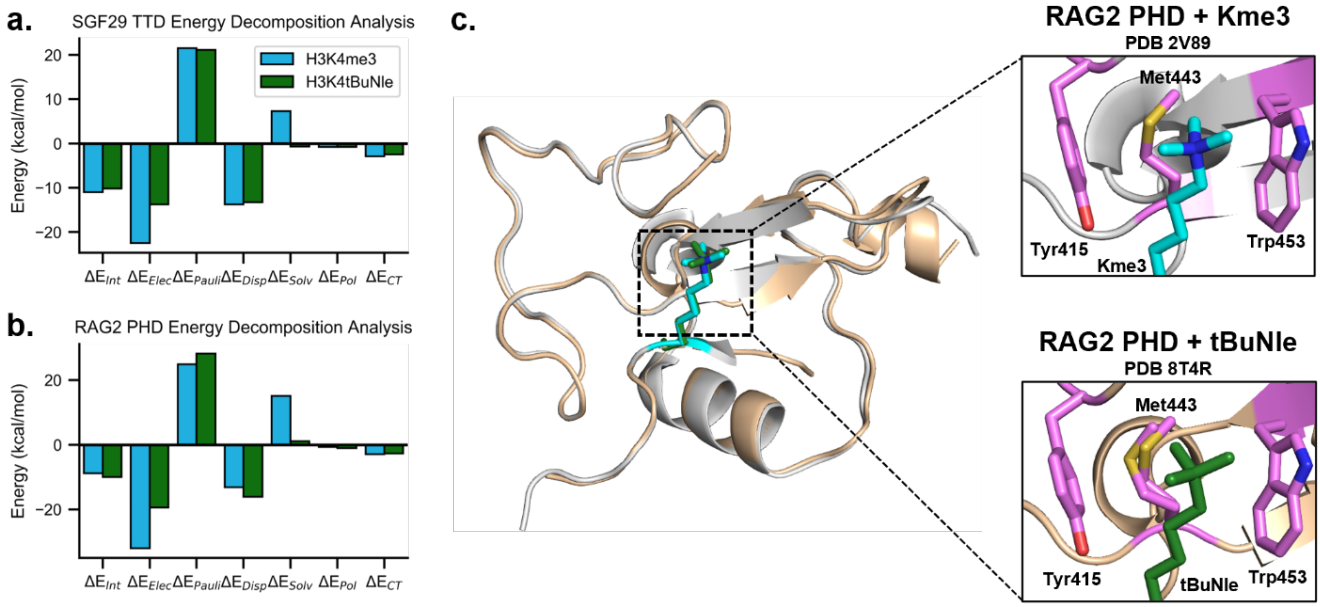



**Figure 3.** Bar graph of energy decomposition analysis calculations for (a) the SGF29 TTD and (b) RAG2 PHD binding H3K4me3 and H3K4tBuNle calculated at the M06-2X/6-311++G(d,p), SMD(Water) level of theory. (c) Overlay of crystal structures of RAG2 PHD binding to H3K4me3 (cyan, PDB 2V89) and H3K4tBuNle (green, PDB 8T4R).

## CONCLUSION

We demonstrate that the binding of histone Kme3 readers to the neutral tBuNle isostere is widespread and report an unprecedented dual mode of binding for Kme3 and tBuNle using the same aromatic cage. We screened a microarray of methyllysine reader proteins and identified that 10 readers of twenty that were quantified bind with comparable or tighter affinity to the neutral analog relative to Kme3.[37,38] This discovery is unexpected, as there are significant differences in selectivity between proteins with highly similar aromatic cages, proteins from the same family, and proteins which recognize the same sequence on the histone tail. Additionally, this report contrasts long-standing precedent that aromatic cages throughout the proteome preferentially bind charged ligands.[27,31,32,34,35]

Our mechanistic studies demonstrate a unique switch in binding to these two ligands, utilizing electrostatically tunable cation–π interactions to bind Kme3 versus using the hydrophobic effect to recognize tBuNle. These are rare examples of protein binding pockets capable of binding with comparable affinities both a charged and uncharged ligand yet employing different mechanisms. Neither structural nor computational evaluation elucidates a reader's selectivity between charged and neutral ligands, thus illustrating the under-appreciated molecular-level nuance of the aromatic cage motif that cannot be easily predicted. This is an example in which Ockham's Razor, which advocates for simpler explanations over those that are more complex, is not supported; despite the universal nature of aromatic cages in binding Kme3, a wide range of selectivities for Kme3 versus tBuNle is observed without a single structural feature that is responsible.[56] This work demonstrates the challenge of extrapolating findings from one molecular recognition system to others as it shows that small differences in the host or protein can result in significant differences in guest or ligand binding. Moreover, these experimental data are important to inform computational and artificial intelligence models of the binding profiles of aromatic cages, a key structural motif across biology.

This work establishes a new framework for the design of selective inhibitors by exploiting differences in charge-dependence among readers in the same family with highly similar binding pockets, which may address a recurrent challenge in therapeutically targeting Kme3 readers. For example, we discovered just one H3K9me3 reader, the UHRF1 TTD, preferentially recognizes tBuNle, offering a promising starting point for therapeutic development for this significant oncological target.[17,40] Moreover, these findings have broader implications beyond histone Kme3 reader proteins, as evolution has converged on aromatic cages to bind small molecules containing trimethylammonium groups, including acetylcholine, choline, betaine, carnitine, and others;[34,57-59] many of these binding events have been shown to be primarily driven by cation–π interactions.[27-29,31,32,34,35] Since there is no selective pressure for these proteins to preferentially bind a cationic ligand over a neutral analog, the variable ability of different aromatic cages to bind a neutral analog of a trimethylammonium may extend to other classes of proteins. Our findings challenge traditional understanding of the molecular recognition of trimethylammonium groups by aromatic cages, indicating the need to further analyze other classes of proteins that bind these key biological molecules.

## ASSOCIATED CONTENT

**Supporting Information**. Materials and methods, ESI-LCMS of peptides and proteins, microarray results, ITC curves, protein crystallography data, computational calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

**Marcey L. Waters** – Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
ORCID ID: 0000-0002-4917-5755
Email: mlwaters@email.unc.edu

### Authors

**Christopher R. Travis** – Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
ORCID ID: 0000-0003-0673-6978
**Kelsey M. Kean** – Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599^
ORCID ID: 0000-0003-4981-9810
**Katherine I. Albanese** – Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
ORCID ID: 0000-0002-2336-1621
**Hanne C. Henriksen** – Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
**Joseph W. Treacy** – Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095-1569
ORCID ID: 0000-0002-8597-0246
**Elaine Y. Chao** – Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095-1569
**K. N. Houk** – Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095-1569
ORCID ID: 0000-0002-8387-5261

### Present Address

^Department of Chemistry, High Point University, High Point, NC 27268


### Author Contributions

C.R.T., K.M.K., K.I.A., and H.C.H. performed experiments and analyzed data under the supervision of M.L.W. J.W.T. and E.Y.C. performed computational calculations under the supervision of K.N.H. C.R.T. and M.L.W. wrote the manuscript, which was edited and approved by all authors.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT


This work was funded by National Institute of General Medical Sciences of the National Institutes of Health under award numbers R01 GM118499 and R35 GM145227 to M.L.W. and K12-GM000678 to K.M.K, as well as the National Institute of Allergy and Infectious Diseases under award number R01 AI141481 to K.N.H. C.R.T. was supported by the NSF Graduate Research Fellowship Program (GRFP). K.I.A. was supported in part by a Burroughs Wellcome Fellowship. Microarray experiments were conducted in the Protein Array and Analysis Core (PAAC) at the University of Texas MD Anderson Cancer Center, which is supported by a grant from the Cancer Prevention and Research Institute of Texas (CPRIT RP 180804). The authors thank Dr. Mark Bedford (UT MD Anderson) for assistance with microarray experiments. This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by the UCLA Institute for Digital Research and Education's Research Technology Group. ITC experiments were conducted at the Macromolecular Interactions Facility at the University of North Carolina at Chapel Hill, which is supported by the National Cancer Institute of the National Institutes of Health under award number P30CA016086. Crystallography data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. SER-CAT is supported by its member institutions, and equipment grants (S10_RR25528, S10_RR028976 and S10_OD027000) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. The authors thank Dr. Ashutosh Tripathy for assistance with ITC experiments.


## ABBREVIATIONS

Kme3, trimethyllysine; tBuNle, tert-butyl norleucine; H3, histone 3; ESP, electrostatic potential; ITC, isothermal titration calorimetry; EEC, enthalpy-entropy compensation; GCE, genetic code expansion.

## REFERENCES


1 Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology* **14**, 1025-1040 (2007).

2 Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology* **19**, 1218-1227 (2012).

3 Musselman, C. A., Khorasanizadeh, S. & Kutateladze, T. G. Towards understanding methyllysine readout. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1839**, 686-693 (2014).

4 Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).

5 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annual review of genetics* **43**, 559-599 (2009).

6 Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1839**, 627-643 (2014).

7 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).

8 Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nature reviews cancer* **10**, 457-469 (2010).

9 Dawson, M. A., Kouzarides, T. & Huntly, B. J. Targeting epigenetic readers in cancer. *New England Journal of Medicine* **367**, 647-657 (2012).

10 Audia, J. E. & Campbell, R. M. Histone modifications and cancer. *Cold Spring Harbor perspectives in biology* **8**, a019521 (2016).

11 Murn, J. & Shi, Y. The winding path of protein methylation research: milestones and new frontiers. *Nature Reviews Molecular Cell Biology* **18**, 517-527 (2017).

12 Barnash, K. D., James, L. I. & Frye, S. V. Target class drug discovery. *Nature chemical biology* **13**, 1053-1056 (2017).

13 Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25**, 15-30 (2007).

14 Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B. & Denu, J. M. Identification and characterization of small molecule inhibitors of a plant homeodomain finger. *Biochemistry* **51**, 8293-8306 (2012).

15 Zhang, M. Y. *et al.* Covalent labeling of a chromatin reader domain using proximity-reactive cyclic peptides. *Chemical Science* **13**, 6599-6609 (2022).
16 Bae, N. *et al.* Developing Spindlin1 small-molecule inhibitors by using protein microarrays. *Nature chemical biology* **13**, 750-756 (2017).
17 Senisterra, G. *et al.* Discovery of small-molecule antagonists of the H3K9me3 binding to UHRF1 tandem tudor domain. *SLAS DISCOVERY: Advancing Life Sciences R&D* **23**, 930-940 (2018).
18 James, L. I. *et al.* Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nature chemical biology* **9**, 184-191 (2013).
19 Stuckey, J. I. *et al.* A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nature chemical biology* **12**, 180-187 (2016).
20 Engelberg, I. A., Foley, C. A., James, L. I. & Frye, S. V. Improved methods for targeting epigenetic reader domains of acetylated and methylated lysine. *Current opinion in chemical biology* **63**, 132-144 (2021).
21 Bennett, R. L. & Licht, J. D. Targeting epigenetics in cancer. *Annual review of pharmacology and toxicology* **58**, 187-207 (2018).
22 Waybright, J. M. & James, L. I. Getting a handle on chemical probes of chomatin readers. *Future Medicinal Chemistry* **13**, 749-763 (2021).
23 Santiago, C., Nguyen, K. & Schapira, M. Druggability of methyl-lysine binding sites. *Journal of computer-aided molecular design* **25**, 1171-1178 (2011).
24 Milosevich, N. & Hof, F. Chemical inhibitors of epigenetic methyllysine reader proteins. *Biochemistry* **55**, 1570-1583 (2016).
25 James, L. I. & Frye, S. V. Chemical probes for methyl lysine reader domains. *Current opinion in chemical biology* **33**, 135-141 (2016).
26 Ortiz, G., Kutateladze, T. G. & Fujimori, D. G. Chemical tools targeting readers of lysine methylation. *Current Opinion in Chemical Biology* **74**, 102286 (2023).
27 Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proceedings of the National Academy of Sciences* **104**, 11184-11188 (2007).
28 Schärer, K. *et al.* Quantification of cation–π interactions in protein–ligand complexes: crystal-structure analysis of factor Xa bound to a quaternary ammonium ion ligand. *Angewandte Chemie International Edition* **44**, 4400-4404 (2005).
29 Salonen, L. M. *et al.* Cation–π interactions at the active site of factor Xa: dramatic enhancement upon stepwise N-alkylation of ammonium ions. *Angewandte Chemie International Edition* **48**, 811-814 (2009).
30 Ma, J. C. & Dougherty, D. A. The cation− π interaction. *Chemical reviews* **97**, 1303-1324 (1997).
31 Zacharias, N. & Dougherty, D. A. Cation–π interactions in ligand recognition and catalysis. *Trends in pharmacological sciences* **23**, 281-287 (2002).
32 Dougherty, D. A. Cation-π interactions involving aromatic amino acids. *The Journal of nutrition* **137**, 1504S-1508S (2007).
33 Shepodd, T. J., Petti, M. A. & Dougherty, D. A. Molecular recognition in aqueous media: donor-acceptor and ion-dipole interactions produce tight binding for highly soluble guests. *Journal of the American Chemical Society* **110**, 1983-1985 (1988).
34 Beene, D. L. *et al.* Cation− π interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: the anomalous binding properties of nicotine. *Biochemistry* **41**, 10262-10269 (2002).
35 Infield, D. T. *et al.* Cation-π interactions and their functional roles in membrane proteins. *Journal of molecular biology* **433**, 167035 (2021).
36 Li, J., Moumbock, A. F. & Günther, S. Exploring cocrystallized aromatic cage binders to target histone methylation reader proteins. *Journal of Chemical Information and Modeling* **60**, 5225-5233 (2020).
37 Kamps, J. J. *et al.* Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nature Communications* **6**, 1-12 (2015).
38 Albanese, K. I. & Waters, M. L. Contributions of methionine to recognition of trimethyllysine in aromatic cage of PHD domains: implications of polarizability, hydrophobicity, and charge on binding. *Chemical Science* **12**, 8900-8908 (2021).
39 Zhu, Y. *et al.* Water and the Cation− π Interaction. *Journal of the American Chemical Society* **143**, 12397-12403 (2021).
40 Ashraf, W. *et al.* The epigenetic integrator UHRF1: on the road to become a universal biomarker for cancer. *Oncotarget* **8**, 51946 (2017).
41 Baril, S. A. *et al.* Investigation of trimethyllysine binding by the HP1 chromodomain via unnatural amino acid mutagenesis. *Journal of the American Chemical Society* **139**, 17253-17256 (2017).
42 Kean, K. M. *et al.* Systematic Variation of Both the Aromatic Cage and Dialkyllysine via GCE-SAR Reveal Mechanistic Insights in CBX5 Reader Protein Binding. *Journal of Medicinal Chemistry* **65**, 2646-2655 (2022).
43 Cornish-Bowden, A. Enthalpy—entropy compensation: a phantom phenomenon. *Journal of Biosciences* **27**, 121-126 (2002).
44 Zhong, W. *et al.* From ab initio quantum mechanics to molecular neurobiology: a cation–π binding site in the nicotinic receptor. *Proceedings of the National Academy of Sciences* **95**, 12088-12093 (1998).
45 Dougherty, D. A. Unnatural amino acids as probes of protein structure and function. *Current opinion in chemical biology* **4**, 645-652 (2000).
46 Lee, Y.-J. *et al.* Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chemical Communications* **52**, 12606-12609 (2016).
47 Galles, G. D. *et al.* Tuning phenylalanine fluorination to assess aromatic contributions to protein function and stability in cells. *Nature communications* **14**, 59 (2023).
48 Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie International Edition* **42**, 1210-1250 (2003).
49 Krone, M. W. *et al.* More Than π–π–π Stacking: Contribution of Amide− π and CH− π Interactions to Crotonyllysine Binding by the AF9 YEATS Domain. *Journal of the American Chemical Society* **142**, 17048-17056 (2020).
50 Travis, C. R., Francis, D. Y., Williams Jr, D. C. & Waters, M. L. Evaluation of acyllysine isostere interactions with the aromatic pocket of the AF9 YEATS domain. *Protein Science* **32**, e4533 (2023).
51 Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894-1900 (2011).
52 Wheeler, S. E. & Houk, K. Substituent effects in cation/π interactions and electrostatic potentials above the centers of substituted benzenes are due primarily to through-space effects of the substituents. *Journal of the American Chemical Society* **131**, 3126-3127 (2009).
53 Matthews, A. G. *et al.* RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V (D) J recombination. *Nature* **450**, 1106-1110 (2007).
54 Zhou, H.-X. & Pang, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical reviews* **118**, 1691-1741 (2018).
55 Champagne, K. S. & Kutateladze, T. G. Structural insight into histone recognition by the ING PHD fingers. *Current drug targets* **10**, 432-441 (2009).
56 Hoffmann, R., Minkin, V. I. & Carpenter, B. K. Ockham's razor and chemistry. *International Journal for the Philosophy of Chemistry* **3**, 3-28 (1997).
57 Chen, C., Malek, A. A., Wargo, M. J., Hogan, D. A. & Beattie, G. A. The ATP-binding cassette transporter Cbc

(choline/betaine/carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds. *Molecular microbiology* **75**, 29-45 (2010).

58 Horn, C. *et al.* Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from Bacillus subtilis for the compatible solutes glycine betaine and proline betaine. *Journal of molecular biology* **357**, 592-606 (2006).

59 Oswald, C. *et al.* Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from Sinorhizobium meliloti in the liganded and unliganded-closed states. *Journal of Biological Chemistry* **283**, 32848-32859 (2008).