






# Mechanism of biomolecular recognition of trimethyllysine by the fluorinated aromatic cage of KDM5A PHD3 finger

Bas J. G. E. Pieters<sup>1</sup>, Maud H. M. Wuts<sup>1</sup>, Jordi Poater<sup>2</sup>, Kiran Kumar<sup>3</sup>, Paul B. White<sup>1</sup>, Jos J. A. G. Kamps<sup>1</sup>, Woody Sherman<sup>4,5</sup>, Ger J. M. Pruijn <sup>1,6</sup>, Robert S. Paton <sup>3</sup>, Thijs Beuming<sup>4,9</sup>, F. Matthias Bickelhaupt <sup>1,7</sup> & Jasmin Mecinović <sup>1,8</sup> 

The understanding of biomolecular recognition of posttranslationally modified histone proteins is centrally important to the histone code hypothesis. Despite extensive binding and structural studies on the readout of histones, the molecular language by which post-translational modifications on histone proteins are read remains poorly understood. Here we report physical-organic chemistry studies on the recognition of the positively charged trimethyllysine by the electron-rich aromatic cage containing PHD3 finger of KDM5A. The aromatic character of two tryptophan residues that solely constitute the aromatic cage of KDM5A was fine-tuned by the incorporation of fluorine substituents. Our thermodynamic analyses reveal that the wild-type and fluorinated KDM5A PHD3 fingers associate equally well with trimethyllysine. This work demonstrates that the biomolecular recognition of trimethyllysine by fluorinated aromatic cages is associated with weaker cation- $\pi$  interactions that are compensated by the energetically more favourable trimethyllysine-mediated release of high-energy water molecules that occupy the aromatic cage.

<sup>1</sup>Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands. <sup>2</sup>ICREA and Departament de Química Inorgànica i Orgànica & IQTCUB, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain. <sup>3</sup>Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, OX1 3TA Oxford, UK. <sup>4</sup>Schrödinger, Inc., 120 West 45th Street, New York, NY 10036, USA. <sup>5</sup>Silicon Therapeutics, 451 D St., Boston, MA 02210, USA. <sup>6</sup>Radboud Institute for Molecular Life Sciences, Radboud University, Geert Grooteplein Zuid 26-28, 6525 GA Nijmegen, The Netherlands. <sup>7</sup>Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands. <sup>8</sup>Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark. <sup>9</sup>Present address: Latham Biopharm Group 101 Main Street, Suite 1400 Cambridge, MA 02142, USA. ✉email: [mecinovic@sdu.dk](mailto:mecinovic@sdu.dk)

Posttranslational modifications on histone proteins have a profound effect on the structure and function of human chromatin<sup>1–3</sup>. Many covalent modifications have been identified and characterized on histone tails and core histones; methylation, acetylation, phosphorylation and ubiquitination have been known for some time, whereas crotonylation and succinylation, among others, have been discovered more recently<sup>4–7</sup>. Methylated lysine residues can exist in the form of monomethyllysine (Kme), dimethyllysine (Kme2) or trimethyllysine (Kme3), and can lead to gene activation or repression, depending on the histone site and methylation state<sup>8</sup>. Histone lysine methylation is dynamically regulated by three classes of functionally related proteins<sup>9</sup>. The installation of the methyl group(s) from *S*-adenosylmethionine (SAM) onto lysine is catalyzed by histone lysine methyltransferases (KMTs)<sup>10</sup>. The opposite reaction, i.e., the removal of methyl group(s) from methylated lysine, is catalyzed either by flavin-dependent lysine specific demethylases or a larger family of non-heme Fe(II) and 2-oxoglutarate (2OG)-dependent histone lysine demethylases (KDMs)<sup>11</sup>. Recent structural and functional studies have revealed that histones that possess unmethylated and methylated lysine residues can be specifically recognized by a large number of reader domain proteins that differ in the composition of the lysine recognition site<sup>12,13</sup>. Electrostatic interactions and H-bonding appear to be of central importance in the recognition of unmethylated lysines by interacting reader domains (e.g. ADD, BAH, PZP)<sup>13</sup>. Similarly, electrostatic interactions and H-bonding play important roles in the readout of the lower methylation states Kme and Kme2 via a cavity-insertion binding mode (e.g. by 53BP1 tandem tudor domains, MBT domains, ankyrin repeats)<sup>13</sup>. Numerous epigenetic reader proteins, including plant homeodomain (PHD) zinc fingers and members of the Royal superfamily (tandem tudor domain, chromodomain and PWWP domain), have been involved in the recognition of trimethyllysine via the so-called surface-groove binding mode<sup>13</sup>. Despite different folding patterns, these reader proteins have a common feature as they all possess electron-rich aromatic cages, most often comprised of 1–4 side chains of Phe, Tyr and Trp, although some cages also include the negatively charged Asp or Glu residues<sup>13,14</sup>. Comparative binding studies between trimethyllysine and its neutral carba analogue led to the conclusion that the recognition of the positively charged trimethyllysine by the aromatic cage containing readers is predominantly driven by a combination of favourable cation– $\pi$  interactions and the release of high-energy water molecules located inside the aromatic cages<sup>15–17</sup>.

To provide a deeper understanding of the origin of molecular recognition of trimethyllysine by aromatic cages, we report here a complementary physical-organic chemistry approach in which

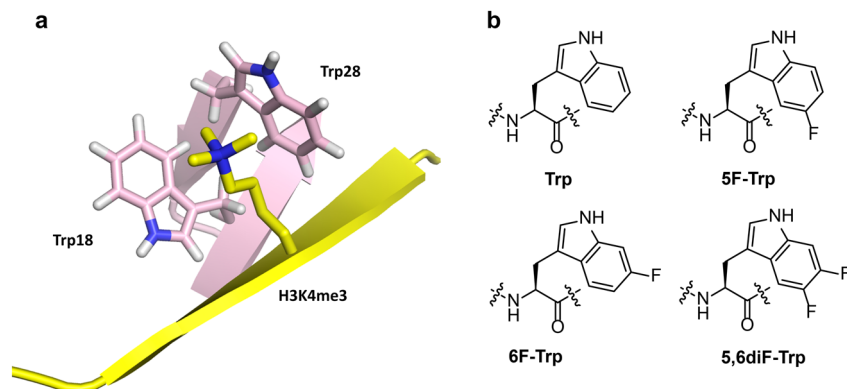
the two electron-rich tryptophans that solely constitute the aromatic cage of KDM5A reader are substituted by fluorinated tryptophans, thus resulting in electron-poorer  $\pi$  systems. Our experimental and computational investigations reveal that despite weaker aromatic character of fluorinated cages, the association between trimethyllysine and fluorinated KDM5A is comparable to that of the wild-type KDM5A. The underlying molecular mechanism for such observation is a well-balanced compensation between energetically less favourable cation– $\pi$  interactions and more favourable release of high-energy water molecules that occupy fluorinated aromatic cages.

## Results

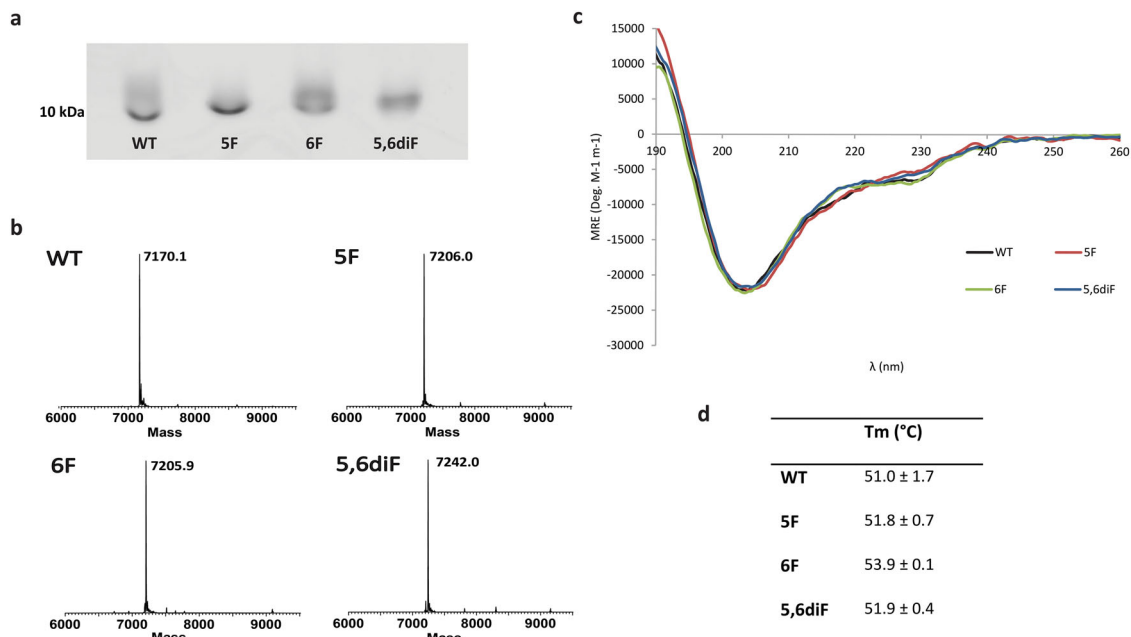
**Physical-organic chemistry approach.** The examination of cation– $\pi$  interactions in biomolecular recognition of positively charged ligands using fluorinated tryptophans was pioneered by Dougherty and coworkers<sup>18–25</sup>. We envisioned that this elegant chemical approach could be employed for probing the involvement of cation– $\pi$  interactions in the readout of trimethyllysine-containing histones by epigenetic reader domains. We chose the PHD3 finger of KDM5A reader protein as a model system, because its specific recognition of H3K4me3 is required for leukemogenesis and, importantly, its aromatic cage is composed of only two tryptophan residues (Trp18 and Trp28) (Fig. 1a)<sup>26</sup>. Since these are the only two tryptophans in the entire KDM5A PHD3 domain, we envisioned that an auxotrophic *E. coli* strain could be used to specifically incorporate fluorinated tryptophan residues directly into its aromatic cage. The presence of only two tryptophans eliminates the risk for perturbation of the reader domain structure as a result of additional fluorinated tryptophan residues outside the region of interest, namely the aromatic cage. This strategy allows us to investigate the effect of the aromatic cage's  $\pi$ -electrons on trimethyllysine recognition by fluorinating the indole rings of the tryptophans at position 5 (5F-Trp), position 6 (6F-Trp), and at positions 5 and 6 (5,6diF-Trp) (Fig. 1b). Fluorination of tryptophan residues was ideal due to (i) fluorine's electronegativity that can be exploited to reduce the electron density of tryptophan's indole rings, and (ii) the comparable size of fluorine and hydrogen, allowing for minimal structural perturbations of the protein. Although it is presently unclear how the fluorination of the aromatic cage affects the energetics of water molecules that occupy such cages, our physical-organic approach also takes into consideration the role of water in the readout of trimethyllysine by the KDM5A PHD3 finger.

## Biochemical and biophysical studies of fluorinated KDM5A.

The KDM5A PHD3 finger was expressed in the auxotrophic, tryptophan deficient, *E. coli* Castellani and Chalmers strain in a



**Fig. 1** The recognition of trimethyllysine by the KDM5A PHD3 finger. **a** View from the complex of the KDM5A PHD3 finger (pink) with H3K4me3 histone peptide (yellow) (PDB: 2KGI); **b** Structures of tryptophan and fluorinated tryptophan residues for probing the strength of cation– $\pi$  interactions.



**Fig. 2 Biochemical and biophysical analyses of wild-type and fluorinated KDM5A PHD3 fingers.** **a** 15% Tris-Tricine SDS-PAGE showing the purified, untagged KDM5A PHD3 fingers. From left to right: Wild-type KDM5A, 5F-KDM5A, 6F-KDM5A and 5,6diF-KDM5A. **b** ESI-MS data showing the masses of the purified KDM5A PHD3 fingers. The wild-type protein has a mass of 7170.1 Da, whereas the 5F and 6F alloproteins show a mass increase of +35.9 and +35.8 Da, respectively, and 5,6diF shows an increase of +71.9 Da. **c** CD spectra showing the purified PHD3 fingers over a range of 190–260 nm. Black: Wild-type KDM5A, red: 5F-KDM5A, green: 6F-KDM5A and blue: 5,6diF-KDM5A. **d** Melting temperatures of KDM5A PHD3 fingers obtained by DSF.

method similar to the procedure described by Budisa and coworkers<sup>27,28</sup>. We successfully produced the protein variants with the three tryptophan analogues: 5F-Trp, 6F-Trp and 5,6diF-Trp in the KDM5A PHD3 reader domain. Additional attempts to incorporate 4,5,6,7tetraF-Trp, the most electron-poor tryptophan analogue, did not lead to production of detectable amounts of fluorinated KDM5A. Interestingly, the wild-type (WT) construct expressed significantly less well than its fluorinated counterparts. Wild-type and fluorinated KDM5A PHD3 fingers were purified using standard biochemical techniques to obtain proteins of high purity in reasonable yields (Fig. 2a and Supplementary Fig. 1). The incorporation of fluorinated tryptophans into KDM5A PHD3 was further verified by denaturing ESI-MS analyses (Fig. 2b). ESI mass spectra confirmed that the wild-type KDM5A PHD3 domain indeed had a mass of 7170.1 Da, and that the incorporation of fluorinated tryptophan residues led to the expected mass increase of 18 Da per fluorine (7206.0, 7205.9 and 7242.0 Da for 5F-KDM5A, 6F-KDM5A and 5,6diF-KDM5A, respectively). Additionally, circular dichroism (CD) spectra indicated that structures of the proteins containing the various fluorinated tryptophan analogues are identical to that of the wild-type protein (Fig. 2c). These results led us to conclude that the KDM5A alloproteins all have similar foldings and that no structural perturbations have been introduced by incorporating fluorinated tryptophans. This finding was further supported by the fact that differential scanning fluorimetry (DSF) experiments showed no decrease in the alloprotein's melting temperature when compared with the wild-type protein. Measured melting temperatures were 51.0 ± 1.7 °C for WT KDM5A; 51.8 ± 0.7 °C for 5F-KDM5A; 53.9 ± 0.1 °C for 6F-KDM5A and 51.9 ± 0.4 °C for 5,6diF-KDM5A (Fig. 2d). Despite the fact that the WT protein expressed markedly less well in the auxotrophic strain when compared with its fluorinated counterparts, CD and ESI-MS experiments conducted on the WT protein expressed in both *E. coli* Rosetta BL21 (DE3)pLysS (hereafter referred to as BL21) and *E. coli* Castellani and Chalmers (hereafter referred to as AUX)

showed that both expression strains produced proteins with identical masses and tertiary structures (Supplementary Figs. 2 and 3). It can therefore be concluded that both expression systems produce the WT KDM5A PHD3 finger with identical structural properties.

#### Thermodynamic analyses of KDM5A–H3K4me3 association.

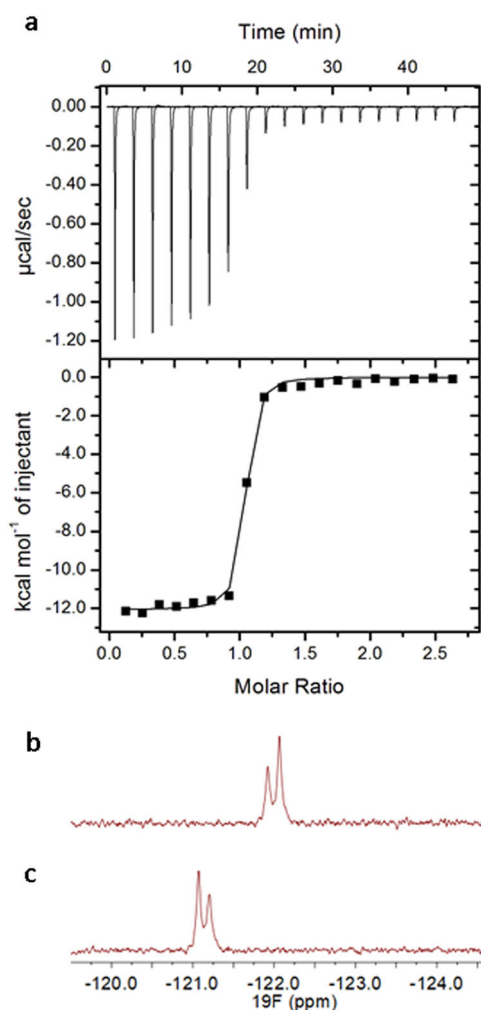
After characterizing the KDM5A proteins, the effect of fluorination of the PHD3 finger of KDM5A on binding to the H3K4me3 peptide (sequence: ARTKme3QTARKS) was examined by isothermal titration calorimetry (ITC), which provided values of the Gibbs free energy of binding ( $\Delta G^\circ$ ), enthalpy of binding ( $\Delta H^\circ$ ) and entropy of binding ( $\Delta S^\circ$ ) (Table 1 and Supplementary Fig. 4). First, a comparison was made between WT protein expressed in the AUX strain and WT protein expressed in the BL21 strain, which had also been used for KDM5A expression in our previous studies<sup>16,29</sup>. The observed binding affinities were indistinguishable, with a  $K_d$  of 54 ± 6 nM for BL21 WT-KDM5A and 48 ± 8 nM for AUX WT-KDM5A (Table 1). Notably, thermodynamic parameters  $\Delta G^\circ$  (BL21-KDM5A:  $-9.9 \pm 0.1$  kcal mol<sup>-1</sup> vs AUX-KDM5A:  $-10.0 \pm 0.1$  kcal mol<sup>-1</sup>),  $\Delta H^\circ$  (BL21-KDM5A:  $-11.6 \pm 0.1$  kcal mol<sup>-1</sup> vs AUX-KDM5A:  $-11.9 \pm 0.1$  kcal mol<sup>-1</sup>) and  $-\Delta S^\circ$  (BL21-KDM5A:  $1.7 \pm 0.2$  kcal mol<sup>-1</sup> vs AUX-KDM5A:  $1.9 \pm 0.1$  kcal mol<sup>-1</sup>) were also observed to be virtually indistinguishable (Table 1). In conjunction with CD and ESI-MS data, these results imply that both WT proteins are identical with respect to their biophysical and binding properties.

Next, we carried out comparative thermodynamic analysis for binding of H3K4me3 with WT KDM5A PHD3 and its fluorinated counterparts. H3K4me3 bound to all four KDM5A reader domain variants with virtually equal binding affinity; the measured  $\Delta G^\circ$  values for all KDM5A–H3K4me3 systems were observed to be  $-9.9 \pm 0.1$  kcal mol<sup>-1</sup> (Table 1). The examination of the enthalpic and entropic terms of  $\Delta G^\circ$ , furthermore, showed that although small differences between WT and fluorinated

**Table 1** Thermodynamic data for binding of H3K4me3 to KDM5A PHD3.

	$K_d$ (nM)	$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$\Delta H^\circ$ (kcal mol <sup>-1</sup> )	$-\Delta\Delta S^\circ$ (kcal mol <sup>-1</sup> )
BL21 WT-KDM5A	54 ± 6	-9.9 ± 0.1	-11.6 ± 0.1	1.7 ± 0.2
AUX WT-KDM5A <sup>a</sup>	48 ± 8	-10.0 ± 0.1	-11.9 ± 0.1	1.9 ± 0.1
5F-KDM5A	49 ± 12	-10.0 ± 0.1	-10.9 ± 0.1	0.9 ± 0.3
6F-KDM5A	57 ± 17	-9.9 ± 0.2	-12.1 ± 0.1	2.2 ± 0.3
5,6diF-KDM5A	70 ± 15	-9.8 ± 0.1	-11.7 ± 0.2	1.9 ± 0.3

10-mer H3K4me3 peptide (ARTKme3QTARS) was used. Values obtained from 7–9 repeated ITC experiments. BL21 and AUX indicate the expression system used for WT protein expression with BL21 corresponding to the Rosetta BL21 (DE3) pLys5 and AUX to the Castellani and Chalmers auxotrophic strains, respectively. The stoichiometry (H3K4me3:KDM5A,  $n$ ) = 0.97–1.03.  
<sup>a</sup>Measured in replicate.



**Fig. 3** Binding of H3K4me3 to the 6F-KDM5A PHD3 finger. **a** ITC curve of 10-mer H3K4me3 histone peptide binding to 6F-KDM5A; **b** <sup>19</sup>F NMR spectrum of 6F-KDM5A; **c** <sup>19</sup>F NMR spectrum of 6F-KDM5A in the presence of the H3K4me3 peptide.

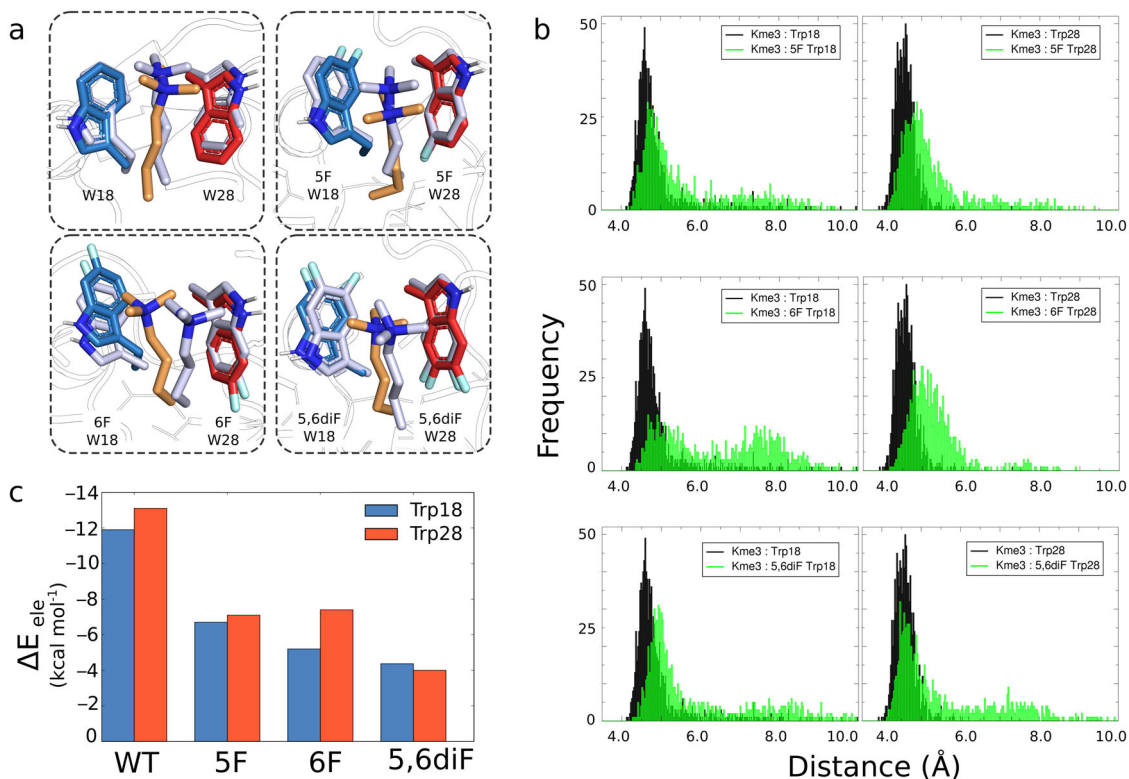
KDM5A are present, they are not significant. 5F-KDM5A showed a decreased enthalpy when compared with wild-type KDM5A, with a  $\Delta\Delta H^\circ$  of  $1.0 \pm 0.1$  kcal mol<sup>-1</sup>. The decrease in enthalpy was, however, completely compensated by an increase in entropy, with a  $-\Delta\Delta S^\circ$  of  $-1.0 \pm 0.3$  kcal mol<sup>-1</sup>. Binding thermodynamics for 6F-KDM5A and wild-type KDM5A are identical within standard error (Fig. 3a, Table 1). Notably, 5,6diF-KDM5A, the most electron-poor aromatic cage in our panel of cages, displayed a very similar thermodynamic signature to wild-type KDM5A. We did not observe any significant differences in values of the free energy of binding, as well as in its enthalpic and

entropic contributions (Table 1). Taken together, our thermodynamic data show that addition of electron-withdrawing fluorine substituents to the indole rings that solely constitute the KDM5A's PHD3 aromatic cage, does not reduce the protein's binding affinity for the positively charged trimethyllysine of H3K4me3. Based on the related examinations of cation- $\pi$  interactions between ammonium cations and tryptophan residues in protein-ligand associations that display significant reduction and linear trend of binding affinity upon fluorination of tryptophan, our studies suggest that cation- $\pi$  interactions are not solely responsible for binding of H3K4me3 to the KDM5A PHD3 finger, as no linear reduction in binding affinity is observed upon increased fluorination of the aromatic cage<sup>19</sup>. As will be discussed below, desolvation effects provide another important contribution to the overall binding process, giving an explanation for maintaining the same binding affinities.

**<sup>19</sup>F NMR studies of fluorinated KDM5A.** We employed <sup>19</sup>F NMR spectroscopy to compare the free and H3K4me3-bound forms of 6F-KDM5A (Fig. 3b, c). The <sup>19</sup>F NMR spectrum of free 6F-KDM5A showed two peaks ( $-121.9$  ppm and  $-122.1$  ppm), thus supporting the presence of two fluorine atoms located at 6F-Trp18 and 6F-Trp28 of the PHD3 finger of KDM5A. Upon binding of H3K4me3 to 6F-KDM5A, we observed down-field shifts of approximately +1 ppm ( $-121.1$  ppm and  $-121.2$  ppm), consistent with the magnitude of shifts found in other ligand binding studies utilizing 6F-Trp labelled proteins<sup>30</sup>. A down-field shift was also observed upon binding of H3K4me3 to 5F-KDM5A (Supplementary Fig. 5). These results indicate that the positively charged trimethyllysine moiety present in the H3K4me3 peptide interacts with the fluorinated tryptophans incorporated into the aromatic cage.

Additional CD spectroscopic analysis confirmed that 5F-KDM5A and 6F-KDM5A remained stable during the NMR measurements performed at 15 °C (Supplementary Figs. 6 and 7). Moreover, CD analyses indicated a small change in the protein's structure upon formation of the 5F-KDM5A-H3K4me3 and 6F-KDM5A-H3K4me3 complex. A small shift in mean residual ellipticity (MRE) between 215–240 nm corresponds to a more extensive  $\beta$ -sheet conformation. This observation is in line with the finding that H3K4me3 peptide forms a third antiparallel  $\beta$ -strand when complexed with the PHD3 domain of KDM5A, as also visible in the reported KDM5A-H3K4me3 structure (Fig. 1a)<sup>26</sup>.

**Molecular dynamics simulations of KDM5A-H3K4me3 complexes.** After experimentally determining that H3K4me3 binds to fluorinated PHD3 fingers of KDM5A, we carried out molecular dynamics (MD) simulations to examine the behaviour over time of the reader-ligand complex and effects of fluorination on key interactions for binding. Four variations of the PHD3 finger of KDM5A were simulated, including the wild-type (PDB: 2KGI) and three variants containing F-substituted Trp18-Trp28



**Fig. 4** Molecular dynamics simulations of wild-type and fluorinated KDM5A PHD3 fingers with H3K4me3. **a** The active site at time 0 ns (light blue) overlaid with the structure observed at the end of the simulation at time 100 ns (H3K4me3, orange; Trp18, blue; Trp28, red) for each system. **b** Histograms of the distance between the H3K4me3 ammonium cation and each aromatic side chain in the Trp18-Trp28 cages. **c** Average  $\Delta E_{\text{ele}}$  values of the ammonium cation of H3K4me3 to each aromatic side chain in the Trp18-Trp28 cages.

aromatic cages: 5F-Trp18/5F-Trp28, 6F-Trp18/6F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28. Adopting a recently described molecular mechanics-based approach<sup>29,31</sup>, the four systems were solvated in a 10 Å truncated octahedral box of TIP3P water<sup>32</sup>, neutralised explicitly with either sodium or chloride ions, and simulated for 100 ns using the Amberff12SB force field.

In all cases of fluorinated KDM5A, the trimethyllysine side chain of H3K4me3 occupies the aromatic cage throughout the simulation (Fig. 4a). Flexibility of the H3 chain to prioritize this interaction is demonstrated for the mutated systems (Supplementary Fig. 8). For KDM5A containing 5F-Trp18/5F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28 behaviour of terminal H3 residues shows great flexibility (Supplementary Fig. 8b and d), compared with the wild-type simulation that shows little difference in the H3 backbone geometry (Supplementary Fig. 8a).

Non-covalent cation- $\pi$  interactions are formed between H3K4me3 and both Trp residues, where we define this using an established geometric cutoff of 6 Å (Fig. 4b and Supplementary Figs. 9–11)<sup>33</sup>. To quantify the strength of these energetically favourable cation- $\pi$  interactions, average  $\Delta E_{\text{ele}}$  values were calculated between the quaternary ammonium cations of H3K4me3 to each aromatic side chain of Trp18-Trp28 (Fig. 4c and Supplementary Table 1). Effects on  $\Delta E_{\text{ele}}$  from fluorination on the Trp side chain suggest a general trend  $\text{WT} > 5\text{F} \geq 6\text{F} > 5,6\text{diF}$  when comparing just the indole heavy atoms or with inclusion of the electronegative fluorine substituents (Supplementary Table 2). These results indicate that the positively charged trimethyllysine predominantly interacts with the  $\pi$ -system of the aromatic cage, and that possible interaction with the electronegative fluorine substituents does not significantly contribute to the stabilization<sup>34</sup>. Fluorination results in less favourable electrostatic contributions to cation- $\pi$  interactions<sup>35</sup>,

consistent with our findings from quantum chemical studies and energy decomposition analyses (see below). Binding of H3K4me3 to 6F-KDM5A leads to differences in  $\Delta E_{\text{ele}}$  values when comparing 6F-Trp18 and 6F-Trp28 (Supplementary Table 2) and agrees with the bimodal distribution of distance between the cation and  $\pi$ -face (Supplementary Fig. 8b). An overall preference for Trp28 over Trp18 is also observed for the systems, except for KDM5A containing 5,6diF-Trp where this interaction is almost equal (Fig. 4c). The stronger interaction between the quaternary ammonium cation with Trp28 has been previously observed for D-Kme3, trimethylornithine and trimethylhomolysine<sup>29,31</sup>. We also examined the distance calculated from the N<sup>+</sup> atom of H3K4me3 to the 5- and 6-membered rings of the Trp18/Trp28, 5F-Trp18/5F-Trp28, 6F-Trp18/6F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28 side chains (Supplementary Figs. 12 and 13). At time 100 ns, virtually no difference is observed when comparing the distance of the cation to the pyrrole or benzene substructure for both wild-type and F-substituted Trp18-Trp28 side chains. In line with this observation, also our quantum chemical calculations reveal only minor changes.

#### Quantum chemical analyses in the gas and aqueous phase.

Next, we aimed to elucidate the nature of the non-covalent interactions between the Kme3 side-chain of the histone peptide and the aromatic cage that consists of two fluorinated tryptophan residues of the KDM5A PHD3 finger (hereafter designated as TRP2 fragment) to understand the underlying origin of the recognition. We characterized quantum chemically the energetics and bonding mechanism in the four model complexes, using dispersion-corrected density functional theory at BLYP-D3BJ/TZ2P and COSMO for simulating aqueous solution, as

**Table 2 Quantum-chemical bonding analysis in TRP2-Kme3 systems.**

	TRP2-Kme3 <sup>a</sup>	5F-TRP2-Kme3 <sup>b</sup>	6F-TRP2-Kme3 <sup>b</sup>	5,6diF-TRP2-Kme3 <sup>b</sup>
$\Delta E(\text{aq})$	-10.2	-10.4	-10.3	-10.3
$\Delta E(\text{aq})_{\text{strain}}$	0.1	0.3	0.3	0.3
$\Delta E(\text{aq})_{\text{int}}$	-10.3	-10.7	-10.6	-10.6
$\Delta E(\text{desolv})_{\text{int}}$	17.3	14.5	14.0	11.7
$\Delta E_{\text{int}}$	-27.6	-25.2	-24.6	-22.3
$\Delta E_{\text{Pauli}}$	20.8	20.4	21.0	19.8
$\Delta V_{\text{elstat}}$	-15.0	-12.2	-12.0	-9.0
$\Delta E_{\text{oi}}$	-13.0	-12.8	-13.0	-12.6
$\Delta E_{\text{disp}}$	-20.4	-20.5	-20.5	-20.5
$d(\text{H}_{\text{Me}}-\text{C}_{\text{TRP-6MR}})$	3.38	3.50	3.54	3.52
$d(\text{H}_{\text{Me}}-\text{C}_{\text{TRP-5MR}})$	2.78	2.70	2.68	2.70

Computed at BLYP-D3BJ/TZ2P with COSMO to simulate aqueous solution. Energies in kcal mol<sup>-1</sup>, distances in Å. Structural rigidity imposed by the protein backbone is simulated through constrained geometry optimizations. See also Eqs. (1-3) in the "Methods" section.

<sup>a</sup>TRP2 frozen, Kme3 entirely free.

<sup>b</sup>TRP2 frozen,  $\alpha$ -methyl carbon fixed to position in TRP2-Kme3 optimization.

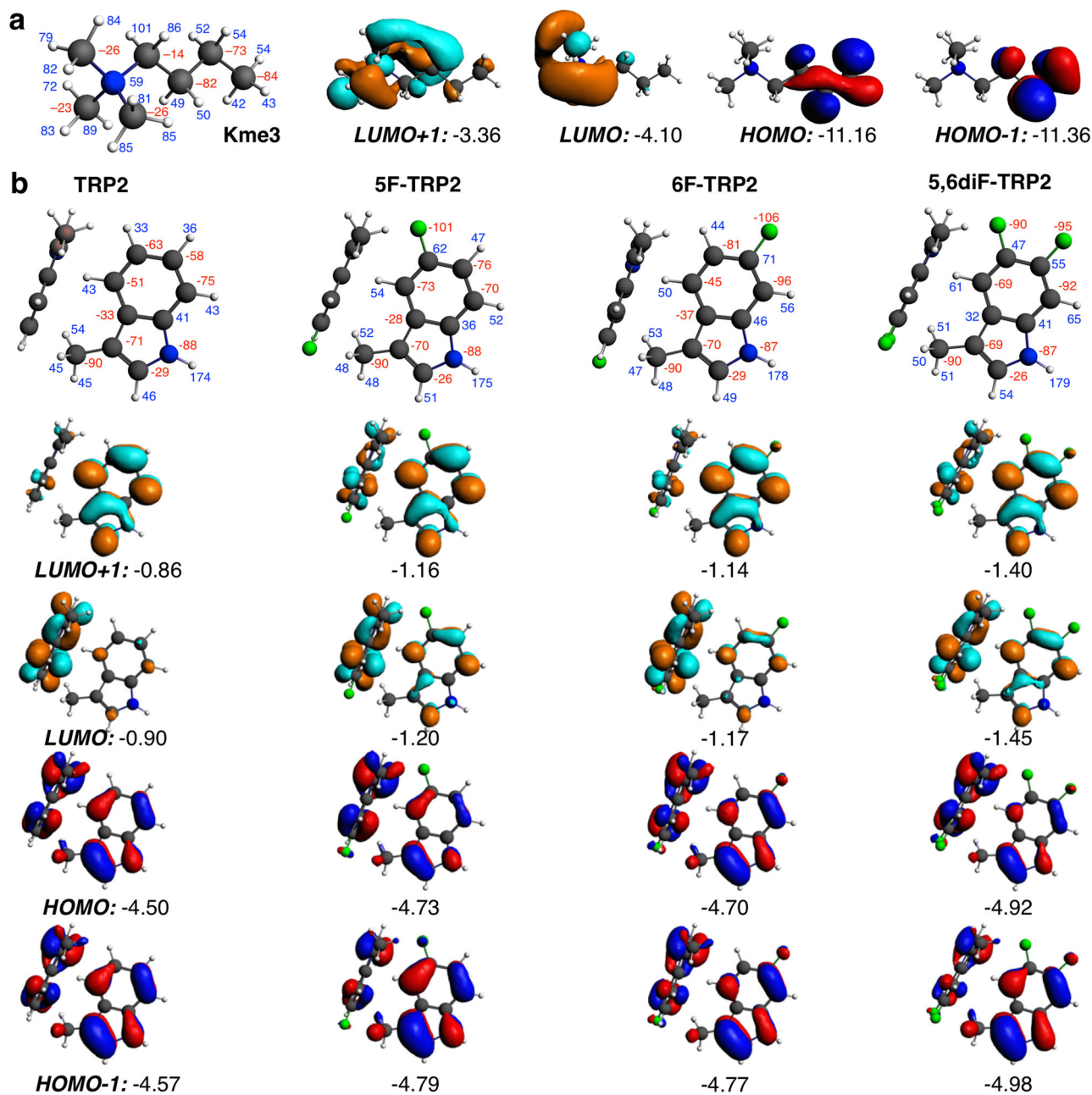
implemented in the ADF program (Supplementary Table 3)<sup>36</sup>. Recently, this methodology has been successfully employed in the quantum chemical exploration and bonding analyses of trimethyllysine analogues by TRP2<sup>16,31</sup>. As previously discussed, from Table 2, it is seen how TRP2-Kme3 presents a bond energy of -10.2 kcal mol<sup>-1</sup>. This energy is almost identical with the corresponding instantaneous interaction energy  $\Delta E(\text{aq})_{\text{int}}$  of -10.3 kcal mol<sup>-1</sup>, due to an almost negligible deformation strain energy ( $\Delta E(\text{aq})_{\text{strain}} = 0.1$  kcal mol<sup>-1</sup>) associated with the subtle variation of geometry upon complexation. The interaction energy between Kme3 and TRP2 without water is stronger ( $\Delta E_{\text{int}} = -27.6$  kcal mol<sup>-1</sup>), because the system presents an unfavourable desolvation energy of 17.3 kcal mol<sup>-1</sup>. The interaction energy in the absence of water,  $\Delta E_{\text{int}}$ , can be decomposed into Pauli repulsion ( $\Delta E_{\text{Pauli}} = 20.8$  kcal mol<sup>-1</sup>), electrostatic attraction ( $\Delta V_{\text{elstat}} = -15.0$  kcal mol<sup>-1</sup>), orbital interaction ( $\Delta E_{\text{oi}} = -13.0$  kcal mol<sup>-1</sup>) and dispersion ( $\Delta E_{\text{disp}} = -20.4$  kcal mol<sup>-1</sup>) terms.

Furthermore, we performed an analogous series of analyses as the one described above, but this time for di- and tetrafluorinated TRP2 as the aromatic cage with Kme3. First, bond energies hardly change for difluorinated 5F-TRP2-Kme3 and 6F-TRP2-Kme3, or tetrafluorinated 5,6diF-TRP2-Kme3 systems ( $\Delta E(\text{aq}) = -10.3$ – $-10.4$  kcal mol<sup>-1</sup>). The same happens for both the deformation strain and interaction energies, with a maximum change of 0.1 kcal mol<sup>-1</sup>. So, even with the presence of fluorine on TRP2, complexation only very slightly changes the geometry of Kme3 side chain. However, changes appear when these interactions are analysed without water. Fluorination of TRP2 causes a weakening of the interaction in the absence of water,  $\Delta E_{\text{int}}$ , between Kme3 and TRP2 of 2.4 and 3.0 kcal mol<sup>-1</sup> for 5F-TRP2-Kme3 and 6F-TRP2-Kme3, respectively, and of 5.3 kcal mol<sup>-1</sup> for 5,6diF-TRP2-Kme3 (Table 2). The weakening in  $\Delta E_{\text{int}}$  upon fluorination is countered by a less unfavourable desolvation energy. The larger desolvation energy of TRP2-Kme3 can be associated with the removal of solvent around the positive charge of the Kme3 side chain ammonium group. In the case of the fluorinated systems, obviously, the same desolvation of Kme3 still occurs. The fact that the electronegative fluorine atoms pull charge out of the aromatic rings reduces the desolvation energy of the latter, which leads to the computed overall less unfavourable  $\Delta E(\text{desolv})_{\text{int}}$  values.

The observation that  $\Delta E_{\text{int}}$  in the gas phase weakens from -27.6 to -25.2, -24.6, and -22.3 kcal mol<sup>-1</sup> for TRP2-Kme3, 5F-TRP2-Kme3, 6F-TRP2-Kme3, and 5,6diF-TRP2-Kme3, respectively, led us to additionally carry out the energy decomposition analysis of the interaction energy. First, it is observed that aforementioned weakening is not the result of the

Pauli repulsion term, which remains quite constant among the complexes ( $\Delta E_{\text{Pauli}} = 19.8$ – $21.0$  kcal mol<sup>-1</sup>), with a maximum difference of 1.0 kcal mol<sup>-1</sup> compared with unfluorinated TRP2 system. This constant Pauli term is in agreement with the minor geometrical changes among the different systems, which can be also followed from distances enclosed in Table 2. The closest H-C distances between an NMe<sub>3</sub><sup>+</sup> H atom and a C atom of a tryptophan in TRP2-Kme3 is 2.78 Å, while the same H atom is 3.38 Å away from the closest C atom of the other tryptophan (Supplementary Fig. 14). For the fluorinated systems, the former distance is slightly shortened (2.68–2.70 Å), whereas the latter is lengthened (3.50–3.54 Å). Distances between the quaternary N atom of Kme3 and the centroids of the five- and six-membered rings of TRP2 can be found in Supplementary Table 4.

We find that the trend in the interaction energy  $\Delta E_{\text{int}}$  originates from the electrostatic attraction  $\Delta V_{\text{elstat}}$ . This attraction is less favourable by 2.8–3.0 kcal mol<sup>-1</sup> for difluorinated, and by 6.0 kcal mol<sup>-1</sup> for tetrafluorinated TRP2 when compared with the TRP2 cage, a trend that we attribute to weaker cation- $\pi$  interactions (Table 2). The weakening of the electrostatic potential is caused by the fact that the electronegative fluorine substituents pull electronic charge density away from the aromatic core (Fig. 5), thus reducing the quadrupole of the rings. This is clearly observed by comparison of the two extreme systems, TRP2-Kme3 and 5,6diF-TRP2-Kme3, that present the strongest effect. In the former, only one carbon in the six-membered ring acquires a net positive partial charge, whereas in the latter, four such partially positively charged carbon atoms exist (Fig. 5). The more positively charged carbon atoms in the six-membered ring reduce the quadrupole, which causes a less favourable electrostatic interaction with positively charged trimethyllysine. The effect is less pronounced for the disubstituted 5F-TRP2-Kme3 and 6F-TRP2-Kme3 systems that have two positively charged carbon atoms in the ring skeleton. Finally, the same constant behaviour as observed for  $\Delta E_{\text{Pauli}}$  also applies to the orbital interaction term  $\Delta E_{\text{oi}}$ , with a maximum difference of 0.4 kcal mol<sup>-1</sup> with fluorinated TRP2 cages (Table 2). The frontier orbitals involved in the interaction between Kme3 and TRP2 are depicted in Fig. 5 for both fragments. The incorporation of fluorine substituents onto tryptophan residues does not affect the shape of the corresponding frontier orbitals; in particular, the interaction between the donor orbitals of TRP2 and the acceptor orbitals of Kme3 is not altered. This finding is further supported by the overlap between the  $\pi$  orbitals of TRP2 and the acceptor orbitals of Kme3 (Supplementary Table 5), with very close values among the four different systems under analysis. The same constant

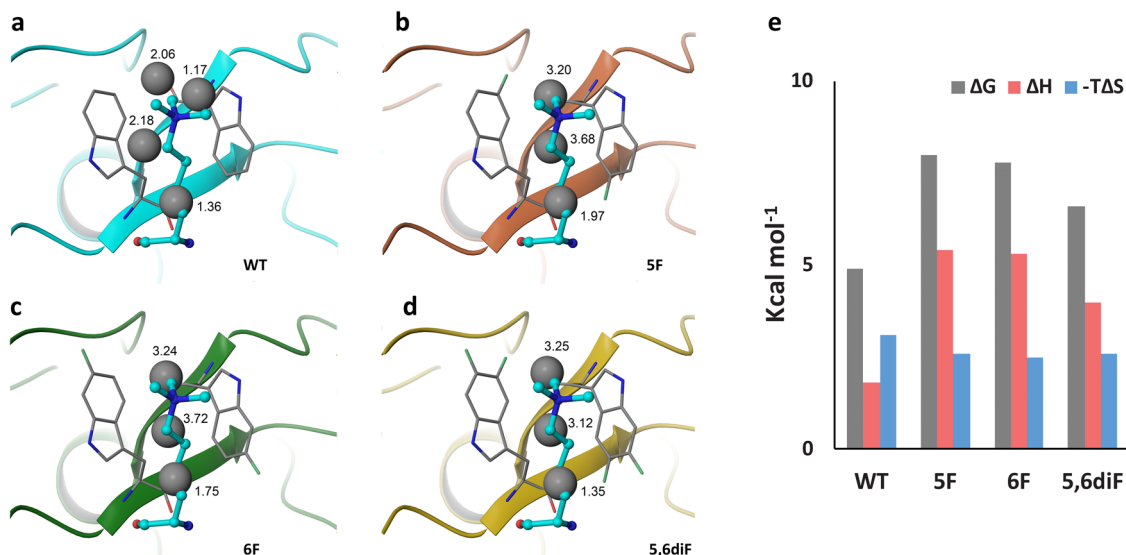


**Fig. 5** Quantum chemical analysis of TRP2-Kme3 interactions. VDD atomic charges (in milli-a.u., red = negative, blue = positive), and frontier orbitals (with orbital energies in eV, isosurface drawn at 0.03) of (a) Kme3 and (b) TRP2, 5F-TRP2, 6F-TRP2 and 5,6diF-TRP2, computed at BLYP-D3BJ/TZ2P.

behaviour is also displayed by the dispersion correction term  $\Delta E_{\text{disp}}$ , which undergoes a negligible change of  $0.1 \text{ kcal mol}^{-1}$  upon fluorination. It is noteworthy that the  $\Delta E_{\text{disp}}$  term contributes the largest to the interaction, however, it has no effect on trends because of its relatively constant value (Table 2).

Further insight into the effect of fluorination of the aromatic cage on the interaction with Kme3 can be gained by estimating the interaction of the cationic nitrogen of Kme3 with either exclusively the five-membered ring or exclusively the six-membered ring of TRP2. We achieved this by introducing tailor-made, for this purpose, modifications into our model systems. Thus, we have constructed modifications of our TRP2-Kme3 system by just keeping one five-membered ring of one TRP unit and one six-membered ring of the other TRP unit, whereas Kme3 has been simplified to  $\text{NMe}_4^+$  (the same procedure has been applied to the fluorinated systems,

Supplementary Fig. 15). Next, we have calculated the energy change  $\Delta E$  associated with the model isodesmic reaction for equilibrium between the  $\text{NMe}_4^+$ -6-membered ring and  $\text{NMe}_4^+$ -5-membered ring (Supplementary Fig. 15).  $\Delta E$  amounts to  $-1.30$ ,  $-1.52$ ,  $-1.48$  and  $-1.77 \text{ kcal mol}^{-1}$  for unfluorinated, 5-mono-fluorinated, 6-mono-fluorinated, and 5,6-difluorinated systems, respectively, all computed at the same BLYP-D3BJ/TZ2P with COSMO level (Supplementary Table 6). These values reveal that the interaction of the  $\text{NMe}_4^+$  cation is more favourable with the five-membered rings than with the six-membered rings by  $1.3$ – $1.8 \text{ kcal mol}^{-1}$ , with a larger difference in case of fluorinated rings. The EDA analyses performed on these systems show that the more favourable interaction of the cation with the five-membered rings is due to an accordingly more favourable electrostatic interaction between the  $\text{NMe}_4^+$  and the same, in all four systems (Supplementary Table 6). This electrostatic



**Fig. 6** Water thermodynamic calculations for the solvation of the aromatic cage of KDM5A PHD3 fingers. **a** Wild-type KDM5A; **b** 5F-KDM5A; **c** 6F-KDM5A; **d** 5,6diF-KDM5A; **e** Thermodynamic parameters for the solvation of the aromatic cages of four KDM5A proteins used in this study.

preference goes with a shorter distance between the  $\text{NMe}_4^+$  and the five-membered ring (Table 2) together with the fact that the five-membered ring is more negatively charged than the six-membered ring (336 vs. 285 milli-a.u., Supplementary Fig. 16). Furthermore, the electrostatic term  $\Delta V_{\text{elstat}}$  is even less favourable in case of the fluorinated systems because the carbon atoms of the six-membered ring bonded to the F atoms become positively charged, thus interacting less favourably with the positively charged H atoms of  $\text{NMe}_4^+$  (Supplementary Fig. 16). On the other hand, the electrostatic interaction between the  $\text{NMe}_4^+$  and the five-membered ring is hardly affected by fluorination. This is in line with the fact that the five-membered rings in TRP2 are more remote from the fluorine substituents and undergo only slight changes in its atomic charges. We recall that, in the model systems discussed above, we have used, for consistency, the same distances between  $\text{NMe}_4^+$  and the six- and five-membered rings as in the full TRP2 model systems. We stress however that we arrive at the same trends and conclusions if we allow for full geometrical relaxation in these further simplified model systems. Just for comparison, the equivalent isodesmic reaction energies in that case are  $-1.00$ ,  $-1.41$  and  $-1.79$  kcal mol<sup>-1</sup> for the unfluorinated, monofluorinated and difluorinated simplified model system, respectively (note that 5- and 6-substitution now lead to one and the same equilibrium geometry). Thus overall, we can conclude that the five-membered rings of TRP2 contribute more to binding to Kme3, and even more so in case of fluorinated aromatic cages.

#### Water thermodynamic analysis of fluorinated aromatic cages.

Water thermodynamic computations, which combine MD simulations with statistical thermodynamic analysis of water molecules, provided strong evidence that desolvation of aromatic cages of trimethyllysine-binding reader proteins is energetically favourable process<sup>16</sup>. We conceived that fluorination of tryptophan residues that constitute the aromatic cage of the KDM5A PHD3 finger presumably leads to altered energetics of high-energy water molecules in their proximity. Therefore, water thermodynamic analyses were carried out to compute thermodynamic parameters for water molecules located in wild-type and fluorinated KDM5A (Fig. 6, Supplementary Table 7 and Supplementary Figs. 17 and 18). For wild-type KDM5A, four high-energy hydration sites were identified, whereas KDM5A PHD3

fingers that possess fluorinated tryptophan residues have three hydration sites. Despite having one water molecule fewer, fluorinated KDM5A displayed a more unfavourable free energy of solvation. The total free energy contributions from desolvation were calculated to be  $-4.9$  kcal mol<sup>-1</sup> for WT KDM5A, and  $-8.0$ ,  $-7.8$ , and  $-6.6$  kcal mol<sup>-1</sup> for 5F-KDM5A, 6F-KDM5A, and 5,6diF-KDM5A, respectively (Fig. 6 and Supplementary Table 7). The increase in the free energy of solvation appears to be a result of more unfavourable enthalpy of solvation. This finding implies that fluorination of the aromatic cage results in a more favourable free energy change upon displacement of water molecules by Kme3 binding. These results support the quantum chemical analysis of the KDM5A-H3K4me3 association, as these computations predicted a compensation mechanism due to a less favourable electrostatics term and a more favourable desolvation term (Table 2). It should be noted, however, that the quantum chemically computed trend of increasingly more favourable desolvation upon 5,6-difluorination was not fully reflected by the water thermodynamic calculations, suggesting that additional energetic factors may be involved in the binding process. For example, the water thermodynamic calculations are based on a molecular mechanics force field that neglects quantum mechanical effects, whereas the quantum mechanical calculations neglect dynamic and entropic information. Despite this fact, the water thermodynamic calculations support the general conclusion that more favourable desolvation upon fluorination of the tryptophan residues constituting the KDM5A's PHD3 aromatic cage compensates for the less favourable interactions of trimethyllysine with the weakened quadrupole of the aromatic cage.

#### Discussion

Understanding the molecular origin of biomolecular recognition processes that play essential roles in human health and disease is important from a basic molecular perspective as well as from a biomedical perspective. Despite extensive examinations of non-covalent interactions in various chemical and biological systems in the past two decades<sup>37–39</sup>, our understanding of the underlying mechanisms that drive biomolecular recognition is partly understood at best, and among others, leads to continual difficulties in rational design of drugs that specifically bind protein targets<sup>40,41</sup>. The phenomenon of biomolecular recognition is further complicated by incomplete understanding of the role of water in binding



processes, although recent computational efforts, in particular, have made significant advances in understanding the structure and energetics of water in protein binding pockets<sup>42–44</sup>. Our work highlights that cooperative experimental and computational investigations enable the examination of the recognition of trimethyllysine-containing histones by epigenetic reader KDM5A at the unprecedented level of detail. Employing a physical-organic chemistry approach allowed us to evaluate the three key contributors that dictate the readout of trimethyllysine: (i) Solute–solute interactions, i.e., cation– $\pi$  interactions between the positively charged trimethyllysine and the electron-rich aromatic cage of the KDM5A PHD3 finger; (ii) Ligand desolvation, i.e., partial desolvation of trimethyllysine upon the KDM5A–Kme3 complex formation; and (iii) Protein desolvation, i.e., desolvation of the aromatic cage of KDM5A upon Kme3 binding. A strategy in which the aromatic character of tryptophan residues is perturbed by the introduction of fluorine substituents, while keeping all other parameters of the KDM5A–H3K4me3 system unaltered, eliminates the contribution from trimethyllysine desolvation in our comparative analyses (as this energetically unfavourable term is present in all systems). Our thermodynamic results (Table 1) showing that H3K4me3 interacts equally well with the electron-rich aromatic cage of wild-type KDM5A PHD3 and comparatively electron-poorer aromatic cages of fluorinated KDM5A, are markedly different to binding studies of the related protein–ligand systems; it has commonly been observed that binding of cations by fluorinated tryptophan or phenylalanine residues is governed by significantly weaker cation– $\pi$  interactions<sup>19,22,45,46</sup>. Our MD simulations and quantum chemical analyses support these findings, by providing evidence that H3K4me3 binding to fluorinated aromatic cages of the KDM5A PHD3 finger or fluorinated TRP2 fragments is associated with an electrostatic weakening of cation– $\pi$  interactions when compared with wild-type KDM5A/TRP2. Notably, the water thermodynamic calculations on the PHD3 finger of KDM5A that possesses tryptophan or its fluorinated counterparts reveal that the energetics of water molecules that occupy aromatic cages is altered upon fluorination of the tryptophan residues. While 3–4 high-energy water molecules are present inside all aromatic cages, the free energy of solvation is more unfavourable in aromatic cages comprised of fluorinated tryptophan residues; these results are line with an increased hydrophobicity of fluorinated benzene relative to benzene<sup>47</sup>. Collectively, our thermodynamic binding studies and computational analyses reveal that the association between the positively charged trimethyllysine and F-substituted tryptophan residues that constitute the aromatic cage of the PHD3 domain of KDM5A is maintained by weaker cation– $\pi$  interactions (when compared with the wild-type aromatic cage) that are compensated by energetically more favourable desolvation of aromatic cages (when compared with the wild-type aromatic cage) upon trimethyllysine binding. More detailed examinations of biomolecular recognition of histones will greatly contribute to our basic understanding of the histone code<sup>48</sup>, which postulates that the molecular landscape of posttranslational modifications on histone proteins is tightly associated with interactions with chromatin-associated proteins, thus altering the chromatin structure and function.

This work demonstrates that a holistic physical-organic chemistry approach, based on synergistic experimental and computational tools, enables a more advanced understanding of biomolecular recognition of trimethyllysine-containing histones by epigenetic reader proteins. It is envisioned that compelling physical-organic chemistry approaches, which collectively examine non-covalent interactions and desolvation effects, along with modern chemical biology approaches<sup>49–52</sup> will importantly contribute to a better understanding of underlying molecular mechanisms that

govern the specific recognition of other types of posttranslational modifications found on histones and other proteins.

## Methods

**Synthesis of 5,6-difluorotryptophan.** Supplementary Fig. 19 shows the schematic presentation of the synthetic protocol for the preparation of 5,6-difluorotryptophan. A suspension of 5,6-difluoroindole (501 mg, 3.27 mmol, 1 equiv.) and L-serine (688 mg, 6.54 mmol, 2 equiv.) in AcOH and Ac<sub>2</sub>O (18 mL, 5:1) was heated to 70 °C under Ar atmosphere in microwave vial. After 16 hours of stirring the solvent was coevaporated with toluene. Crude brown oil was purified by column chromatography (MeOH in CH<sub>2</sub>Cl<sub>2</sub> (0–5%) and with AcOH (0.1%)), affording *N*-acetyl-5,6-difluorotryptophan (680 mg, 2.41 mmol, 74%) as a yellowish oil. <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$ : 7.36–7.29 (m, 1H), 7.17–7.11 (m, 1H), 7.11 (d, *J* = 4.0 Hz, 1H), 4.65 (dt, *J* = 8.0, 5.0 Hz, 1H), 3.23 (ddd, *J* = 15.0, 8.0, 0.5 Hz, 1H), 3.08 (ddd, *J* = 14.8, 8.0, 0.5 Hz, 1H), 1.89 (s, 3H). <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD)  $\delta$ : 173.5, 171.7, 131.5 (d, *J*<sub>C-F</sub> = 21.5 Hz), 124.7 (d, *J* = 3.5 Hz), 122.8 (d, *J*<sub>C-F</sub> = 7.5 Hz), 110.2 (d, *J*<sub>C-F</sub> = 4.5 Hz), 104.3 (d, *J* = 18.5 Hz), 98.4 (d, *J*<sub>C-F</sub> = 21.5 Hz), 53.2, 20.9. ESI-MS calcd for C<sub>13</sub>H<sub>13</sub>F<sub>2</sub>O<sub>3</sub>N<sub>2</sub> [M + 1]<sup>+</sup> 283.0894, found 283.0888. A solution of *N*-acetyl-5,6-difluorotryptophan (485 mg, 1.71 mmol) was dissolved in aqueous NaOH (25 mL, 4 M) and heated at 100 °C for 16 h. The reaction mixture was then cooled and acidified to pH 1. Solvent was evaporated and the crude product was purified by preparative HPLC, affording racemic 5,6-difluorotryptophan as a TFA salt (140 mg, 0.39 mmol, 23%) as a yellowish oil. <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$ : 7.44–7.36 (m, 1H), 7.26–7.17 (m, 2H), 4.21 (dd, *J* = 7.0, 5.0 Hz, 1H), 3.39 (ddd, *J* = 15.5, 5.0, 0.5 Hz, 1H), 3.30 (ddd, *J* = 15.5, 5.0, 0.5 Hz, 1H). <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD)  $\delta$ : 170.2, 148.1 (dd, *J*<sub>C-F</sub> = 168.0, 15.5 Hz), 145.7 (dd, *J*<sub>C-F</sub> = 165.0, 15.5 Hz), 131.9 (d, *J*<sub>C-F</sub> = 10.5 Hz), 126.0 (d, *J*<sub>C-F</sub> = 3.5 Hz), 122.3 (d, *J*<sub>C-F</sub> = 7.0 Hz), 106.9 (d, *J*<sub>C-F</sub> = 5.5 Hz), 104.3 (d, *J*<sub>C-F</sub> = 19.5 Hz), 98.8 (d, *J*<sub>C-F</sub> = 21.5 Hz), 52.9, 25.9. <sup>19</sup>F NMR (377 MHz, MeOD)  $\delta$  –78.2 (s, 3F), –148.3 (m, 1F), –151.6 (m, 1F). ESI-MS calcd for C<sub>11</sub>H<sub>11</sub>F<sub>2</sub>O<sub>2</sub>N<sub>2</sub> [M + 1]<sup>+</sup> 241.0789, found 241.0797.

**Auxotrophic production of KDM5A.** The wild-type KDM5A PHD3 finger (Homo sapiens, uniprot ID: P29375, residues 1598–1663) fused to GST was expressed in Rosetta BL21 (DE3)pLysS *E. coli* containing the KDM5A-GST construct in TB medium supplemented with the appropriate antibiotics. At OD<sub>600</sub> ~0.6, expression was induced with 0.4 mM IPTG and 0.1 mM ZnCl<sub>2</sub> (final concentration) and cultured overnight at 16 °C. Cells were then harvested, lysed and purified using GST affinity. The GST tag was cleaved off with TEV-protease under reducing conditions (10 mM dithiothreitol), and the KDM5A PHD3 finger was subsequently purified by size exclusion chromatography on a Superdex 75 column using 20 mM TRIS-HCl pH 7.5, 50 mM NaCl, 1 mM DTT as running buffer. Protein concentration was measured spectrophotometrically using a Denovix DS-11 spectrophotometer and protein masses were confirmed by ESI-MS analyses. Wild-type and fluorinated PHD3 finger of KDM5A-GST (Homo sapiens, uniprot ID: P29375, residues 1598–1663) expressed in the auxotrophic *E. coli* (Migula) Castellani and Chalmers strain were cultured in either New Minimal Medium (NMM) or in Unnatural amino acid New Minimal Medium (UNMM) supplemented with appropriate antibiotics, respectively. NMM was prepared as described by Budisa and coworkers<sup>27,28</sup>. In brief, NMM contained 100 mM K<sub>2</sub>HPO<sub>4</sub>, 55 mM KH<sub>2</sub>PO<sub>4</sub>, 20 mM D-glucose, 8.5 mM NaCl, 7.5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM MgSO<sub>4</sub>, 10 mg l<sup>-1</sup> biotin, 10 mg l<sup>-1</sup> Thiamine-HCl, 1 mg l<sup>-1</sup> CaCl<sub>2</sub> and FeCl<sub>3</sub>, 1  $\mu$ g l<sup>-1</sup> CuSO<sub>4</sub>, MnCl<sub>2</sub>, ZnCl<sub>2</sub>, NaMoO<sub>4</sub> and 50 mg l<sup>-1</sup> of each individual amino acid. UNMM was prepared similarly except that tryptophan was substituted by the desired fluorinated tryptophan analogue, at a final concentration of 25 mg l<sup>-1</sup>. *E. coli* (Migula) Castellani and Chalmers containing the wild-type KDM5A construct was cultured in NMM at 37 °C. At OD<sub>600</sub> ~0.6, the NMM medium was refreshed by harvesting the cells, after which they were resuspended in fresh NMM. Expression was then induced with 0.1 mM IPTG and 0.1 mM ZnCl<sub>2</sub> (final concentrations). The cells were subsequently cultured for 3 h at 37 °C, after which the culture was harvested, lysed and purified as described above. Fluorinated tryptophan analogues were introduced into KDM5A as follows: *E. coli* (Migula) Castellani and Chalmers containing the KDM5A construct was initially cultured in NMM at 37 °C. At OD<sub>600</sub> ~0.6, the cells were harvested and subsequently washed three times with 0.9% NaCl at room temperature. Following the washing steps, the cells were resuspended in fresh UNMM. Expression was then induced with 1.0 mM IPTG and 0.1 mM ZnCl<sub>2</sub> (final concentrations). The cells were subsequently cultured for 3 h at 37 °C after which the culture was harvested, lysed and purified as described above.

**Circular dichroism.** CD experiments were carried out at a protein concentration of 0.1 mg ml<sup>-1</sup> in 10 mM phosphate buffer, pH 7.5, on a J-815 circular dichroism spectropolarimeter. The samples were measured over a range of 180–260 nm with normal sensitivity and a bandwidth of 1 nm. Scanning was performed at 50 nm per minute, a data integration time (D.I.T.) of 0.5 s and a data pitch of 0.5 nm. The spectra for each protein are a result of 10 accumulations.

**Differential scanning fluorimetry.** Protein melt curves were obtained as described by Reinhard et al. using a StepOne-Plus Real-Time PCR system (Applied

Biosystems) and MicroAmp fast optical 96-well reaction plates (Applied Biosystems)<sup>53</sup>. SYPRO-Orange protein gel stain (Invitrogen) was used as a reporter dye, emitting fluorescence in the FAM channel. Total reaction volume was 25  $\mu$ l:20  $\mu$ l buffer (25 mM TRIS-HCl pH 7.5, 50 mM NaCl, 1 mM DTT), 2.5  $\mu$ l of 25  $\mu$ M protein and 2.5  $\mu$ l SYPRO-Orange dye (diluted 1:100 in ddH<sub>2</sub>O). Melt curve data were obtained in triplicate, in a temperature range of 20–95 °C at a stepwise temperature increment of 1 °C min<sup>-1</sup>. Obtained data were analyzed using DSF Analysis v3.0.2 software, designed by Niesen et al. (available via <ftp://ftp.sgc.ox.ac.uk/pub/biophysics/>)<sup>54</sup>.

**Isothermal titration calorimetry.** The same batch of H3K4me3 histone peptide (ARTKme3QTARKS, 380  $\mu$ M) was titrated to all KDM5A PHD3 fingers (28  $\mu$ M). Due to lower expression of the auxotrophic WT-KDM5A, H3K4me3 (190  $\mu$ M) and AUX WT-KDM5A (21  $\mu$ M) were used. The buffer used for ITC experiments was the same as the elution buffer used for SEC; 20 mM TRIS-HCl pH 7.5, 50 mM NaCl, 1 mM DTT. Each ITC titration consisted of 19 injections. ITC experiments were performed on the fully automated Microcal Auto-iTC200 (GE Healthcare Life Sciences, USA). Heats of dilution for histone peptides were determined in control experiments, and were subtracted from the titration binding data before curve fitting. Curve fitting was performed by Origin 6.0 (Microcal Inc., USA) using one set of sites binding model. With the exception of the auxotrophic WT-KDM5A-H3K4me3 (replicate), 7–9 independent ITC experiments were carried out for other four reader-histone systems.

**<sup>19</sup>F NMR spectroscopy.** Measurements were obtained on a Bruker AVANCE III 400 MHz system equipped with a BBFO probe capable of <sup>19</sup>F nucleus detection with <sup>1</sup>H decoupling. Samples were prepared using 5 mm Shigemitsu tubes matched to D<sub>2</sub>O to minimize solvent volume required. <sup>19</sup>F NMR experiments were performed 10 mM H<sub>2</sub>KPO<sub>4</sub> pH 7.5, at a concentration of 450  $\mu$ M of 5F-KDM5A/6F-KDM5A and 1 mM of H3K4me3 peptide (ARTKme3QTARKS). All measurements were performed at 288 K. After samples were inserted into the magnet, the sample was shimmed using the lock nucleus in D<sub>2</sub>O and a <sup>1</sup>H spectrum was acquired to assess the quality of the shims. The probe was then manually tuned and matched to <sup>19</sup>F to optimize <sup>19</sup>F detection. A 15  $\mu$ s @ 23 Watts 90-degree pulse was used. <sup>19</sup>F{<sup>1</sup>H} spectra were then acquired with the following parameters: NS = 1.5 k-28 k, d1 = 1, aq = 1.09 s, sw = 20.1 ppm and o1p near -120 ppm. <sup>19</sup>F NMR spectra were externally referenced to CFCl<sub>3</sub> using the frequency of residual solvent signal in the <sup>1</sup>H spectrum and the ratio between the <sup>1</sup>H and <sup>19</sup>F gyromagnetic ratios.

**MD simulations.** Four MD simulations were carried out for 100 ns each using the Amberff12SB force field. A PDB structure for the model representing KDM5A PHD3 (PDB: 2KGI) was used as a template for building the readerKme3 systems. KDM5A residues Trp18 and Trp28 were manually modified to generate the 5F-Trp18/5F-Trp28, 6F-Trp18/6F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28 complexes. Hydrogen atom addition was performed with LEaP. Systems were solvated in a 10 Å truncated octahedral box of TIP3P<sup>32</sup> water and neutralized explicitly with either sodium or chloride counterions. Non-bonding parameters of Zn(II) previously established from studies of KDM4A<sup>35</sup> were employed. Atomic partial charges for 5F-Trp, 6F-Trp, and 5,6diF-Trp correspond to the Restrainted Electrostatic Potential (RESP)<sup>55</sup> charges, as shown in Supplementary Tables 8–10. Parameters for Kme3 were taken from previous work<sup>29</sup>. The final systems were minimized for 1000 cycles of steepest-descent minimization followed by 1000 cycles of conjugate-gradient minimization to remove close van der Waals contacts using the sander program in AMBER12. Equilibration was achieved using PMEMD to heat the systems to 310 K followed by independent MD simulations performed with a periodic boundary condition at a constant pressure of 1 atm with isotropic molecule-based scaling at a time step of 2.0 fs. All simulations used a dielectric constant of 1.0, Particle Mesh Ewald summation<sup>56</sup> to calculate long-range electrostatic interactions and bond-length constraints applied to all bonds to H atoms. Trajectories were saved at 20 ps intervals and visualized using VMD<sup>57</sup>. Electrostatic energies between the terminal modified Kme3 side chain (e)-N atom and the  $\pi$ -system of surrounding aromatic cages were calculated with the NAMDEnergy Plugin<sup>57</sup>. The  $\pi$ -system was defined for tryptophan and fluorinated tryptophans as the side chain indole ring (non-H) atoms. Energy values were measured every 20 ps and averaged over 100 ns.

**Quantum chemical analyses.** All calculations were carried out with the Amsterdam Density Functional (ADF) program using dispersion-corrected density functional theory at the BLYP-D3BJ/TZ2P level of theory<sup>36,58</sup>. The effect of solvation was simulated by means of the Conductor-like Screening Model (COSMO) of solvation as implemented in ADF. The approach has been benchmarked against highly correlated post-Hartree-Fock methods and experimental data and was found to work reliably<sup>59–62</sup>.

The bonding mechanism in our model complexes have been further analysed using quantitative (Kohn-Sham) molecular orbital (MO) theory in combination with an energy decomposition analysis (EDA)<sup>63,64</sup>. The bond energy in aqueous solution  $\Delta E(\text{aq})$  consists of two major components, namely, the strain energy  $\Delta E(\text{aq})_{\text{strain}}$  associated with deforming the Kme3 and the reader from their own equilibrium structure to the geometry they adopt in the complex,

plus the interaction energy  $\Delta E(\text{aq})_{\text{int}}$  between these deformed solutes in the complex (Eq. 1):

$$\Delta E(\text{aq}) = \Delta E(\text{aq})_{\text{strain}} + \Delta E(\text{aq})_{\text{int}} \quad (1)$$

To arrive at an understanding of the importance of desolvation phenomena during the complexation process, we separate the solute-solute interaction  $\Delta E(\text{aq})_{\text{int}}$  into the effect caused by the change in solvation  $\Delta E(\text{desolv})$  and the remaining intrinsic interaction  $\Delta E_{\text{int}}$  between the unsolvated fragments in vacuum  $\Delta E_{\text{int}}$ :

$$\Delta E(\text{aq})_{\text{int}} = \Delta E(\text{desolv})_{\text{int}} + \Delta E_{\text{int}} \quad (2)$$

In the EDA, the intrinsic interaction energy  $\Delta E_{\text{int}}$  can be further decomposed as shown in Eq. 3:

$$\Delta E_{\text{int}} = \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} + \Delta E_{\text{disp}} \quad (3)$$

Here,  $\Delta V_{\text{elstat}}$  corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the deformed fragments which is usually attractive. The Pauli repulsion  $\Delta E_{\text{Pauli}}$  comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsions. The orbital interaction  $\Delta E_{\text{oi}}$  accounts for charge transfer (donor-acceptor interactions between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO-LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment due to the presence of another fragment). Finally, the  $\Delta E_{\text{disp}}$  term accounts for the dispersion interactions based on Grimme's DFT-D3BJ correction. Furthermore, the charge distribution has been analysed using the Voronoi deformation density (VDD) method<sup>65</sup>.

**Water thermodynamic calculations.** Water thermodynamic calculations were performed with the program WaterMap, as described in previous reports<sup>66,67</sup>. All calculations were run in with default settings. In brief, a 2 ns molecular dynamic (MD) simulation of the KDM5A PHD3 finger with the histone peptide removed, was performed using the Desmond molecular dynamic engine with the OPLS2.1 force field<sup>43</sup>. Protein atoms were constrained throughout the simulation. Water molecules from the simulation were then clustered into hydration sites for thermodynamic analysis. Enthalpy values for each hydration site were obtained by computing the average non-bonded interaction for each water molecule in the cluster over the course of the MD simulation. Entropy values were calculated using a numerical integration of a local expansion of the entropy in terms of spatial and orientational correlation functions<sup>68,69</sup>. The contribution of water-free energy to the binding free energy of the peptide was approximated by the sum of the free energies of hydration sites displaced by the ligand upon binding.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information file. Other relevant data are available from the corresponding author upon reasonable request.

Received: 24 October 2019; Accepted: 6 May 2020;

Published online: 01 June 2020

## References

- Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Tan, M. et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016–1028 (2011).
- Farrelly, L. A. et al. Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535–539 (2019).
- Galligan, J. J. et al. Methylglyoxal-derived posttranslational arginine modifications are abundant histone marks. *Proc. Natl Acad. Sci. USA* **115**, 9228–9233 (2018).
- Zhang, D. et al. Metabolic regulation of gene expression by histone lactylation. *Nature* **574**, 575–580 (2019).
- Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
- Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
- Qian, C. & Zhou, M.-M. SET domain protein lysine methyltransferases: structure, specificity and catalysis. *Cell. Mol. Life Sci.* **63**, 2755–2763 (2006).

11. Cloos, P. A. C., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140 (2008).
12. Andrews, F. H., Strahl, B. D. & Kutateladze, T. G. Insights into newly discovered marks and readers of epigenetic information. *Nat. Chem. Biol.* **12**, 662 (2016).
13. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
14. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21**, 564–578 (2011).
15. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl Acad. Sci. USA* **104**, 11184–11188 (2007).
16. Kamps, J. J. A. G. et al. Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nat. Commun.* **6**, 8911 (2015).
17. Pieters, B. J. G. E. et al. Installation of trimethyllysine analogs on intact histones via cysteine alkylation. *Bioconjugate Chem.* **30**, 952–958 (2019).
18. Dougherty, D. A. Cation- $\pi$  interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163–168 (1996).
19. Dougherty, D. A. The cation- $\pi$  interaction. *Acc. Chem. Res.* **46**, 885–893 (2013).
20. Gallivan, J. P. & Dougherty, D. A. Cation- $\pi$  interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
21. Ma, J. C. & Dougherty, D. A. The Cation- $\pi$  Interaction. *Chem. Rev.* **97**, 1303–1324 (1997).
22. Xiu, X., Puskas, N. L., Shanata, J. A. P., Lester, H. A. & Dougherty, D. A. Nicotine binding to brain receptors requires a strong cation- $\pi$  interaction. *Nature* **458**, 534 (2009).
23. Cashin, A. L., Petersson, E. J., Lester, H. A. & Dougherty, D. A. Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **127**, 350–356 (2005).
24. Dougherty, D. A. Physical organic chemistry on the brain. *J. Org. Chem.* **73**, 3667–3673 (2008).
25. Tavares, X. D. S. et al. Variations in binding among several agonists at two stoichiometries of the neuronal,  $\alpha 4\beta 2$  nicotinic receptor. *J. Am. Chem. Soc.* **134**, 11474–11480 (2012).
26. Wang, G. G. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847–851 (2009).
27. Budisa, N. et al. Proteins with  $\beta$ -(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Sci.* **10**, 1281–1292 (2001).
28. Minks, C., Huber, R., Moroder, L. & Budisa, N. Atomic mutations at the single tryptophan residue of human recombinant annexin V: effects on structure, stability, and activity. *Biochemistry* **38**, 10649–10659 (1999).
29. Belle, R. et al. Investigating d-lysine stereochemistry for epigenetic methylation, demethylation and recognition. *Chem. Commun.* **53**, 13264–13267 (2017).
30. Arntson, K. E. & Pomerantz, W. C. K. Protein-observed fluorine NMR: a bioorthogonal approach for small molecule discovery. *J. Med. Chem.* **59**, 5158–5171 (2016).
31. Al Temimi, A. H. K. et al. Recognition of shorter and longer trimethyllysine analogues by epigenetic reader proteins. *Chem. Commun.* **54**, 2409–2412 (2018).
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
33. Kumar, K. et al. Cation- $\pi$  interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **9**, 2655–2665 (2018).
34. Wheeler, S. E. & Houk, K. N. Substituent effects in cation/ $\pi$  interactions and electrostatic potentials above the centers of substituted benzenes are due primarily to through-space effects of the substituents. *J. Am. Chem. Soc.* **131**, 3126–3127 (2009).
35. Cortopassi, W. A., Kumar, K. & Paton, R. S. Cation- $\pi$  interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity. *Org. Biomol. Chem.* **14**, 10926–10938 (2016).
36. te Velde, G. et al. Chemistry with ADF. *J. Comput. Chem.* **22**, 931–967 (2001).
37. Persch, E., Dumele, O. & Diederich, F. Molecular recognition in chemical and biological systems. *Angew. Chem. Int. Ed.* **54**, 3290–3327 (2015).
38. Salonen, L. M., Ellermann, M. & Diederich, F. Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed.* **50**, 4808–4842 (2011).
39. Cockroft, S. L. & Hunter, C. A. Chemical double-mutant cycles: dissecting non-covalent interactions. *Chem. Soc. Rev.* **36**, 172–188 (2007).
40. Whitesides, G. M. & Krishnamurthy, V. M. Designing ligands to bind proteins. *Q. Rev. Biophys.* **38**, 385–395 (2005).
41. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.* **14**, 95 (2015).
42. Snyder, P. W., Lockett, M. R., Moustakas, D. T. & Whitesides, G. M. Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Spec. Top.* **223**, 853–891 (2014).
43. Wang, L. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).
44. Wang, L., Berne, B. J. & Friesner, R. A. Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl Acad. Sci. USA* **108**, 1326–1330 (2011).
45. Baril, S. A. et al. Investigation of trimethyllysine binding by the HP1 chromodomain via unnatural amino acid mutagenesis. *J. Am. Chem. Soc.* **139**, 17253–17256 (2017).
46. Lee, Y.-J. et al. Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun.* **52**, 12606–12609 (2016).
47. Kumar, A. & Patwari, G. N. Hydration of fluorobenzenes: a molecular dynamics simulation investigation. *J. Indian Inst. Sci.* **100**, 221–230 (2020).
48. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
49. Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nat. Chem. Biol.* **8**, 417–427 (2012).
50. Müller, M. M. & Muir, T. W. Histones: at the crossroads of peptide and protein chemistry. *Chem. Rev.* **115**, 2296–2349 (2015).
51. David, Y. & Muir, T. W. Emerging chemistry strategies for engineering native chromatin. *J. Am. Chem. Soc.* **139**, 9090–9096 (2017).
52. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Curr. Opin. Chem. Biol.* **45**, 35–47 (2018).
53. Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. Optimization of protein buffer cocktails using Thermofluor. *Acta Crystallogr. Sect. F* **69**, 209–214 (2013).
54. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212 (2007).
55. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP method. *J. Phys. Chem.* **97**, 10269–10280 (1993).
56. Wang, H., Fang, J. & Gao, X. The optimal particle-mesh interpolation basis. *J. Chem. Phys.* **147**, 124107 (2017).
57. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
58. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
59. Fonseca Guerra, C., van der Wijst, T., Poater, J., Swart, M. & Bickelhaupt, F. M. Adenine versus guanine quartets in aqueous solution: dispersion-corrected DFT study on the differences in  $\pi$ -stacking and hydrogen-bonding behavior. *Theor. Chem. Acc.* **125**, 245–252 (2010).
60. Padiál, J. S., de Gelder, R., Fonseca Guerra, C., Bickelhaupt, F. M. & Meciniović, J. Stabilisation of 2,6-diarylpiperidinium cation by through-space polar- $\pi$  interactions. *Chem. Eur. J.* **20**, 6268–6271 (2014).
61. van der Wijst, T., Fonseca Guerra, C., Swart, M., Bickelhaupt, F. M. & Lippert, B. A ditopic ion-pair receptor based on stacked nucleobase quartets. *Angew. Chem. Int. Ed.* **48**, 3285–3287 (2009).
62. Simó Padiál, J. et al. Stabilization of 2,6-diarylanilinium cation by through-space cation- $\pi$  interactions. *J. Org. Chem.* **82**, 9418–9424 (2017).
63. Baerends, E. J., Gritsenko, O. V. & van Meer, R. The Kohn-Sham gap, the fundamental gap and the optical gap: the physical meaning of occupied and virtual Kohn-Sham orbital energies. *Phys. Chem. Chem. Phys.* **15**, 16408–16425 (2013).
64. Bickelhaupt, F. M. & Baerends, E. J. Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry. In: *Rev. Comput. Chem.* Lipkowitz, K. B. & Boyd, D. B., Eds. Wiley-VCH: New York, **15**, 1–86 (2000).
65. Fonseca Guerra, C., Handgraaf, J. W., Baerends, E. J. & Bickelhaupt, F. M. Voronoi deformation density (VDD) charges: assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *J. Comput. Chem.* **25**, 189–210 (2004).
66. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **130**, 2817–2831 (2008).
67. Beuming, T. Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* **80**, 871–883 (2012).
68. Lazaridis, T. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B* **102**, 3531–3541 (1998).
69. Lazaridis, T. Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids. *J. Phys. Chem. B* **102**, 3542–3550 (1998).

## Acknowledgements

We thank the European Research Council (ERC Starting Grant, ChemEpigen-715691, J.M.), the Netherlands Research School for Chemical Biology (NRSCB, J.M.), the Netherlands Organization for Scientific Research (NWO-ALW, NWO-CW, and NWO-EW, F.M.B.), the Generalitat de Catalunya (2017SGR348, J.P.) and the Spanish MINECO (CTQ2016-77558-R and MDM-2017-0767, J.P.) for financial support. K.K. is supported by a World Bank Education Grant. We thank Professor Nediljko Budisa for providing the auxotrophic strain and Professor Jan van Hest for helpful discussions at the early stage of the project.

## Author contributions

J.M. conceived and supervised the project. B.J.G.E.P. and M.H.M.W. expressed and purified proteins. B.J.G.E.P. carried out biophysical and thermodynamic studies, and analysed data. P.B.W. performed NMR studies. J.J.A.G.K. synthesized 5,6-difluorotryptophan. J.P. and F.M.B. carried out quantum chemical analyses and interpreted results. K.K. and R.S.P. carried out MD simulations and analyzed results. W.S. and T.B. carried out water thermodynamic calculations and analyzed results. B.J.G.E.P. and J.M. wrote the manuscript with contributions from W.S., G.J.M.P., R.S.P., T.B. and F.M.B. All authors contributed to editing the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42004-020-0313-2>.

Correspondence and requests for materials should be addressed to J.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020