Biochimica et Biophysica Acta 1850 (2015) 1026–1040

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Molecular dynamics for computational proteomics of methylated histone H3☆

Cédric Grauffel [1], Roland H. Stote, Annick Dejaegere *

Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Centre National de la Recherche Scientifique (CNRS) UMR7104, Université de Strasbourg, 67404 Illkirch, France

ARTICLE INFO

Article history:
Received 10 July 2014
Received in revised form 9 September 2014
Accepted 10 September 2014
Available online 18 September 2014

Keywords:
Molecular dynamics
Protein interaction
Histone
Proteomics
Epigenetics

ABSTRACT

Background: Post-translational modifications of histones, and in particular of their disordered N-terminal tails, play a major role in epigenetic regulation. The identification of proteins and proteic domains that specifically bind modified histones is therefore of paramount importance to understand the molecular mechanisms of epigenetics.
Methods: We performed an energetic analysis using the MM/PBSA method in order to study known complexes between methylated histone H3 and effector domains of the PHD family. We then developed a simple molecular dynamics based predictive model based on our analysis.
Results: We present a thorough validation of our procedure, followed by the computational predictions of new PHD domains specific for binding histone H3 methylated on lysine 4 (K4).
Conclusions: PHD domains recognize methylated K4 on histone H3 in the context of a linear interaction motif (LIM) formed by the first four amino acids of histone H3 as opposed to recognition of a single methylated site. PHD domains with different sequences find chemically equivalent solutions for stabilizing the histone LIM and these can be identified from energetic analysis. This analysis, in turn, allows for the identification of new PHD domains that bind methylated *H3*K4 using information that cannot be retrieved from sequence comparison alone.
General significance: Molecular dynamics simulations can be used to devise computational proteomics protocols that are both easy to implement and interpret, and that yield reliable predictions that compare favorably to and complement experimental proteomics methods. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

Chromatin is the physiologically relevant substrate for all genetic processes inside the nuclei of eukaryotic cells and the regulation of these processes results from dynamic changes in its local and global organization. During the life of the cell, chromatin undergoes various levels of condensation–decondensation, which is essential for processes such as mitosis, replication, repair, recombination and transcription. The basic unit of chromatin is the nucleosome, which is composed of about 146 bp of DNA [1,2] wrapped around an octameric core composed of two copies each of the H2A, H2B, H3 and H4 histone proteins. Histones are small proteins that possess a folded globular domain and a flexible N-terminal tail that protrudes from the nucleosome structure. These N-terminal tails are the targets of different post-translational modifications (PTMs) that include covalent modification of specific amino acids, such as acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines, or ligation of protein domains, like ubiquitination or sumoylation of lysines [3]. These post-translational modifications play active roles in regulating chromatin structure either by directly affecting the interactions between histones and DNA or by triggering the recruitment of effector protein complexes that subsequently act on chromatin. Major efforts are being made to characterize the effect of histone PTMs on gene expression and to decipher their molecular mechanisms of action. PTMs are often linked to specific biological effects, for example, methylation of lysines 4 and 36 of histone H3 (*H3*K4, *H3*K36) is generally associated with transcription activation, while methylation of *H3*K9 and *H3*K27 is linked to transcriptional repression [4–6]. Analysis of these effects led to the development of the histone code model [7] whereby different combinations of histone modifications are linked to specific patterns of gene expression. Furthermore, the discovery of enzymes that not only produce PTMs, but also erase them [3], as well as the discovery of protein domains that recognize similar histone PTMs yet belong to different chromatin regulating complexes underline the complex dynamic and combinatorial nature of histone PTMs [8]. Much current effort is thus aimed at identifying proteins that specifically recognize histones bearing post-translational modifications [9]. Diverse experimental and

☆ This article is part of a Special Issue entitled Recent developments of molecular dynamics.

* Corresponding author.
[1] Present address for C. Grauffel: Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan.

http://dx.doi.org/10.1016/j.bbagen.2014.09.015
0304-4165/

computational proteomics protocols for high-throughput analysis of peptide-binding proteins and protein domains have been developed in the recent years (see for example [10] and references therein) and these methods have been applied for the identification of epigenetic readers. Stable isotope labeling by amino acids in cell culture (SILAC) based mass spectrometry methods are very powerful in this context [11,12] and have been applied on a large scale to identify proteins that bind trimethyl lysine (me3) modifications [13,14]. Peptide arrays [15,16] have formed the basis of large scale studies of histone PTM interactants, while chemical cross linking protocols for PTM reader detection have been developed with the aim of improving the detection of the transient protein–peptide interactions associated with epigenetic regulations [17,18]. Computational proteomics protocols for identifying the recognition motifs of specific protein domains are complementary to these experimental methods and usually consider a combination of both sequence- and structure-based data (see for example [19]). In the peptide–protein domain, such methods have thus far focused primarily on the identification of the ensemble of peptide sequences that can be read by a given protein domain [20]. A recent study, for example, focused on the ensemble of methyl lysine containing peptides in the human proteome that may bind to the CBX6 chromodomain [21]. Plant Homeodomain (PHD) fingers are small conserved domains of ≈60 residues that are binders of modified histone tails. They were first identified in the plant *Arabidopsis thaliana* [22] as a new cysteine-rich domain contained in the homeodomain protein HAT3.1 and have since been found in a variety of nuclear proteins associated with transcription regulation. As binders of modified histone tails, the first PHD fingers that were identified belonged to the BPTF (Bromodomain and Phd domain Transcription Factor) and ING2 (INhibitor of Growth 2) proteins, where they were found to recognize methylation on lysine 4 of histone H3 (*H3*K4) [23–26]. Although initially identified as binders of methylated *H3*K4, a more diverse selectivity in molecular recognition by PHD domains was subsequently revealed. For example, several proteins bind to un-methylated *H3*K4 via their PHD domain (BHC80/PHF21A, TRIM24, the first PHD domain of AIRE and BRPF2, and the first and second domains of CHD4) [27–31]. JARID1C binds methylated *H3*K9 [32] while other PHD-containing proteins, such as Set4 in *Saccharomyces cerevisiae*, have no detectable histone binding activity [16]. The complexity of epigenetic regulation mechanisms is further illustrated by PHD domains that share specific reading of epigenetic marks, but that result in different functional outcomes. For example, in the PHD domains of the ING family, ING3-5 proteins activate transcription through their histone acetylase (HAT) activity, while ING1-2 proteins are histone deacetylases (HDAC) and are thus linked to repression [33–35]. The overall governing factor appears to be related to macromolecular chromatin-binding complex to which the PHD domains belong. Given their role as effectors of epigenetic regulation, PHD fingers have been implicated in a number of diseases, including cancer, immunodeficiency syndromes and neurological disorders [36,37]. It is therefore believed that disruption of PHD binding to histones could, in some cases, be an interesting avenue for therapy [38]. As such, there is a considerable interest in understanding the molecular determinants of PHD binding to histones, as well as identifying potential new histone binders among the PHD domains identified in the human genome. The 3D structures of several PHD fingers in complex with methylated histone peptides have been resolved and provide important insight into their mode of recognition. The PHD finger family possesses a characteristic Cys4-His-Cys3 motif that coordinates two zinc ions and folds into a conserved three-dimensional structure that possesses a central β-sheet (strands β_3 and β_4 on Fig. 1A) and two small α-helices. The mode of recognition of the methylated lysine involves cation-π, as well as hydrophobic and van der Waals contributions [39], although the relative contribution of each interaction to overall affinity is still controversial [40,41]. The selectivity of recognition displayed by several PHD binders for the *H3*K4 methylated side-chain over other methylation sites (such as *H3*K9, *H3*K27, or *H4*K20) is thought to be linked to the sequence context of the methylation mark, although the molecular determinants of selective recognition are not yet fully understood [42]. Spiliotopoulos and co-workers evaluated the total binding energy of unmodified *H3*K4 peptides to autoimmune regulator (AIRE), as well as the effect of alanine point-mutations [43]. Their results highlighted the fact that histone peptide anchoring is mainly driven by van der Waals interactions, which was already observed in the case of methylated histone peptide recognition by the JMJD2A tandem Tudor domain [44]. Proteome-wide analysis of *Saccharomyces Cerevisiae* showed that a large proportion of PHD domains bind methylated histone peptides. Moreover, it was argued that conserved sequence information alone was not sufficient to predict the molecular binding properties of PHD domains [16]. To address the question of prediction, we present a computational protocol based on molecular dynamics simulations, homology modeling and MM-PBSA calculations specifically aimed at identifying protein domains of the PHD family that bind methylated lysine and more precisely, methylated lysine 4 of histone H3. Using a free energy decomposition protocol [45] based on the MM/PBSA method [46], we perform a thorough energetic analysis of complexes between PHD domains and methylated *H3*K4 complexes. This analysis is made possible by the substantial amount of structural information available on complexes between PHD domains and modified histone tails. From this analysis, we developed a simple yet effective computational protocol that allows a genome-wide identification of PHD domains likely to bind specifically to methylated *H3*K4. We then present a thorough validation and application of this protocol.

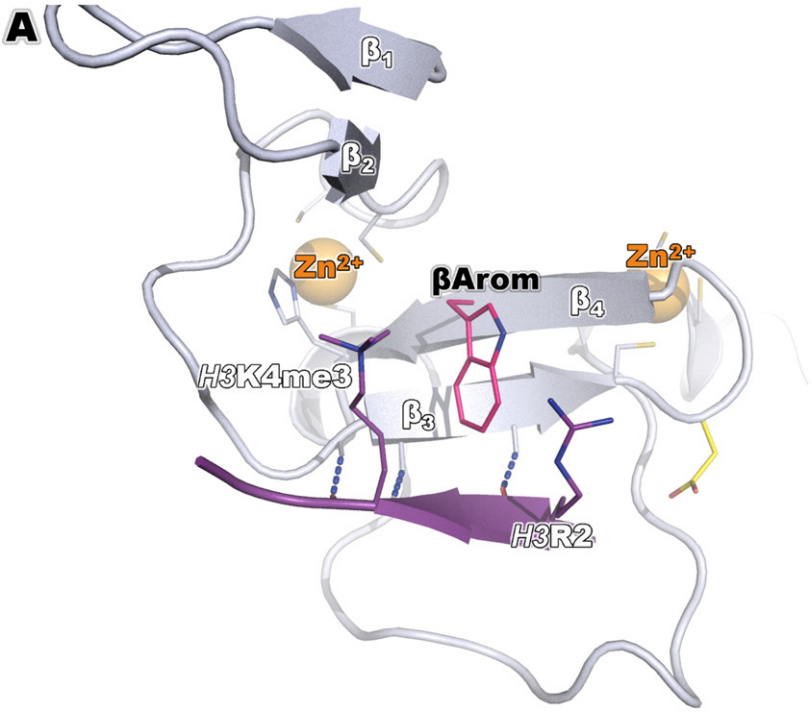
2. Methods

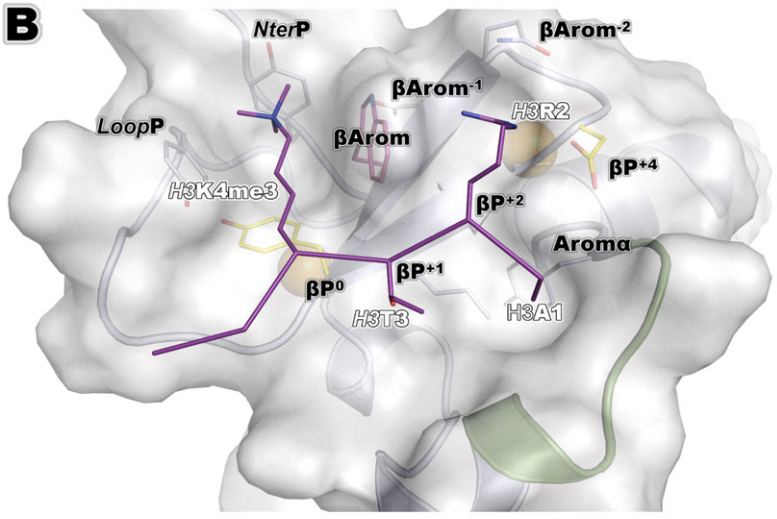
We developed a simple yet effective protocol for predicting the propensity of PHD domains to bind the methylated *H3*K4 of Histone H3 tail. The protocol is schematized in Fig. 2. The method rests on a detailed energetic analysis of the selective recognition of methylated H3 tails in existing structures of PHD–histone peptide complexes using molecular dynamics simulations and a free energy decomposition protocol [45]. From the analysis of these data, a consensus picture emerged of the energetically important interactions that stabilize the complexes. Then, a prediction protocol was devised based on the construction of homology models of PHD domain–histone peptide complexes and on the energetic analysis of these modeled complexes. The different steps of the protocol are detailed below.

2.1. Molecular dynamics simulations of experimental complexes

2.1.1. Preparation of the MD simulations of experimental complexes

Coordinates of PHD domains in complex with methylated *H3*K4 peptides were obtained from the Protein Data Bank (PDB) [55] and are listed in Table 1. For JARID1A in complex with an *H3*K4me3 peptide, there exists both an x-ray and an NMR structure (PDB IDs *3GL6* and *2KGI*, respectively). Due to a partial unfolding of the N-terminal part of the PHD domain in the crystallographic structure, we designed a hybrid structure by taking the coordinates of the first 10 residues of the domain from the NMR structure. The protonation states of all histidine residues at the crystallization pH were determined using the program PROPKA [56]. All other titratable groups were placed in their standard protonation states. Hydrogen atoms were constructed using the HBUILD module of the CHARMM (v33) program [57]. The N- and C-terminal residues of the protein were constructed in their charged state, as was the N-terminal residue *H3*A1 of histone H3 peptide. The C-terminal residue of the histone H3 peptide was constructed with a neutral N-methylamide group. The cysteine residues chelating the zinc ions of PHD domains were constructed in deprotonated forms as would be expected in this environment. The system was then submitted to an energy minimization using harmonic restraints (50 kcal/mol/Å on the backbone and 25 kcal/mol/Å on the side chains, respectively): 2000 steps of energy minimization using the Steepest Descent algorithm was followed by 1000 steps using the conjugated gradient method. The harmonic


A
β1
β2
Zn2+
βArom
Zn2+
β4
H3K4me3
β3
H3R2


B
NterP
βArom-2
βArom-1
LoopP
H3R2
βArom
βP+4
H3K4me3
βP+2
Aromα
βP+1
βP0
H3T3
H3A1

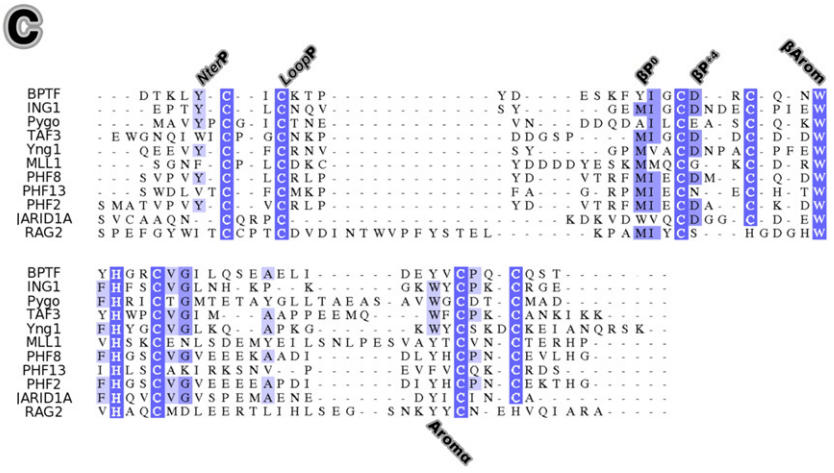

C
NterP
LoopP
βP0
βP+1
βArom
BPTF
ING1
Pygo
TAF3
Yng1
MLL1
PHF8
PHF13
PHF2
JARID1A
RAG2
Aromα

Table 1
Experimental systems studied. List of the proteins that contain a PHD finger domain complexed to a methylated *H3*K4 peptide used in establishing our protocol. The letter *x* stands for the methylation degree of *H3*K4 in the corresponding system. Experimental affinities in μM have been indicated when available (ND = no binding detected). Protein Data Bank IDs are indicated, and NMR structures are labeled with a (*).

Protein name	Ref.	K_d (*H3*K4me$_x$)				PDB structure		
		$x = 0$	$x = 1$	$x = 2$	$x = 3$	ID	*H3*K4	Pept. length
BPTF	[23]			5.0	2.7	*2F6J*	me3	15
ING1	[34]		419	17.3	3.3	*2QIC*	me3	12
ING2	[25]		208	15	1.5	*2G6Q*	me3	12
ING4	[35]		34	9.2	3.0	*2PNX*	me3	12
ING5	[33]		222	16	2.4	*3C6W*	me3	12
JARID1A	[47]		2.8	0.9	0.7	*3GL6*	me3	9
$MLL1_3$	[48]		53	6.9	5.3	*3LQJ*	me3	9
PHF2	[49]					*3KQI*	me3	12
PHF8	[50]				0.95	*3KV4*	me3	24
PHF13						*3O7A*	me3	11
Pygo	[51]	ND	2.2	0.9	1.2	*2VPE*	me2	7
RAG2	[52]			173	34	*2V89*	me3	10
TAF3*	[53]				0.3	*2K17*	me3	13
Yng1*	[54]		50	21	9.1	*2JMJ*	me3	9

restraints were scaled by 0.65 every 500 steps of minimization. The complexes were then solvated in a cubic TIP3P water box of dimensions $65 \times 65 \times 65$ Å^3 [58] and counter-ions (either chloride or sodium) were added to neutralize the charge of the system. Water molecules overlapping the protein, determined by an overlap cutoff of 2.8 Å, were removed.

2.1.2. Molecular dynamics simulations

Molecular dynamics simulations were used to explore the conformational space around the experimental structures. Simulations were performed at a temperature of 300 K using the program NAMD [59] and the CHARMM27 force field [58]. For the modified histone tails, force field parameters developed by us were used [60]. The SHAKE algorithm was used to constrain all bonds between heavy atoms and hydrogens. Non-bonded interactions were truncated at a cutoff of 14 Å, using a switch function for the van der Waals interactions and a shift function for electrostatic interactions [61]. We used four heating phases at 10 K, 100 K, 200 K and 300 K, respectively, followed by an equilibration phase at 300 K for 150 ps. For the production phase, five simulations of 2.5 ns were run using different initial velocity distributions in order to explore conformational space around the crystallographic structure. From each trajectory, only the last 2 ns were used for analysis, resulting in a 10 ns sampling for every experimental system studied. In all simulations, an integration step of 1 fs was used.

2.2. Energetic analysis of the experimental complexes

A free energy decomposition protocol based on the MM/PBSA method [46] was used to obtain a semi-quantitative evaluation of the contribution of all amino-acids from the PHD domains and from the histone peptide to the formation of the complex. In this approach (described in [45]), the free energy is estimated using a thermodynamic cycle and can be expressed as the sum of terms given in Eq. (1)

$$\Delta G_{assoc}^{Protein/Peptide} = \Delta E_{intra} - T\Delta S + \Delta E_{vdW} + \Delta G_{solv}^{Non-polar} + \Delta G_{solv}^{Elec.} + \Delta E_{Elec.}. \quad (1)$$

The solvent contribution to the electrostatics term is calculated using the University of Houston Brownian Dynamics program (UHBD, Release 4.1) [62] with a grid spacing of 0.4 Å, and the intermolecular electrostatics term is calculated using the partial charges in the CHARMM force field. The van der Waals and non-polar contributions are evaluated using the CHARMM program. The non-polar contribution is taken to be proportional to the change of Solvation Accessible Surface Solvent (SASA) scaled by a factor of 0.005 kcal/mol/Å. The free energy is estimated for an ensemble of conformations extracted from the MD simulation, as described in [45]. It has indeed been observed that small structural changes can lead to significant variations in terms of energy in the MM/PBSA method [61]. As the Coulomb energy can fluctuate significantly with protein structural dynamics, this energy is calculated for all conformations, which are then sorted and clustered into twenty-five groups that are affected a weight given their population. For each cluster, the conformation with the energy closest to the cluster average is then selected for the MM/PBSA calculation. Several approximations are introduced in the free energy decomposition, including: *i*) changes in internal energy upon complex formation are neglected, as the structures of the individual unbound protein and peptide used in the MM/PBSA calculation are simply taken from the structure of the complex, and *ii*) changes in conformational entropy are neglected [63]. These approximations result in a semi-quantitative estimation of the binding energy, which can be decomposed into individual contributions of each amino acid of the complex. Despite the approximations involved, free energy decomposition protocols have shown to be very effective in identifying and quantifying important interactions that stabilize protein complexes [63]. This analysis allowed us to identify important recurring interactions in the different complexes, which are described in the Results section.

2.3. Computational protocol for identifying PHDs that specifically bind methylated H3K4

Based on the results of the energetic analysis of the experimental complexes between PHDs and methylated *H3*K4 peptides (following the protocol presented above and based on the data discussed in the Results section), we developed a computational proteomics protocol that combines homology modeling and energetic analysis for identifying PHDs in the human genome specific for methylated *H3*K4. This protocol rests on a bioinformatics filter, followed by homology modeling of complexes that satisfied the sequence requirement, and finally energetic evaluation that requires short molecular dynamics simulations of the modeled complexes to determine whether they would bind methylated *H3*K4me peptides.

2.3.1. Sequence filter and homology modeling of PHD domains

Structural analysis of PHDs suggested the importance of an aromatic cage around methylated H3K4 (*H3*K4me) for specific recognition [42]. Our energetic analysis of the experimental complexes confirmed these observations (see Results section) and highlighted the residue labeled βArom (see Fig. 1B and C) as having a particularly important energetic contribution in stabilizing the methylated *H3*K4. This aromatic amino acid is fully conserved in the sequence alignment of PHDs that recognize methylated lysines. In line with these observations, we first applied a sequence filter to PHDs. The PF00628 sequence database from the PFAM data bank [65] contains sequences of all domains that have been identified as PHD fingers. Using version 21 of the database, we obtained a sequence alignment of all human PHDs. From this alignment, we retained solely those PHDs that displayed an aromatic residue at the position βArom. It turned out that in human sequences, no aromatic other than a Trp is observed at this position. However, the presence of an aromatic residue at this position, while required for binding a methylated lysine, does not guarantee specificity for the N-terminal sequence of H3. Indeed, as detailed above, the PHDs that recognize methylated *H3*K4 also harbor specific residues for recognition of the H3 N-terminal ART sequence. These features, although energetically

Fig. 1. (A) Overview of the anchoring mode of histone peptide on the PHD. The peptide is displayed in violet, and residues *H3*K4me3 and *H3*R2 are shown in sticks. Two residues of the domain are shown in sticks: βArom (magenta) and βP+ 4 (yellow). (B) Name convention for the most important amino acids in the structures of PHD domains. (C) Sequence alignments of the PHD domains of the systems are studied.

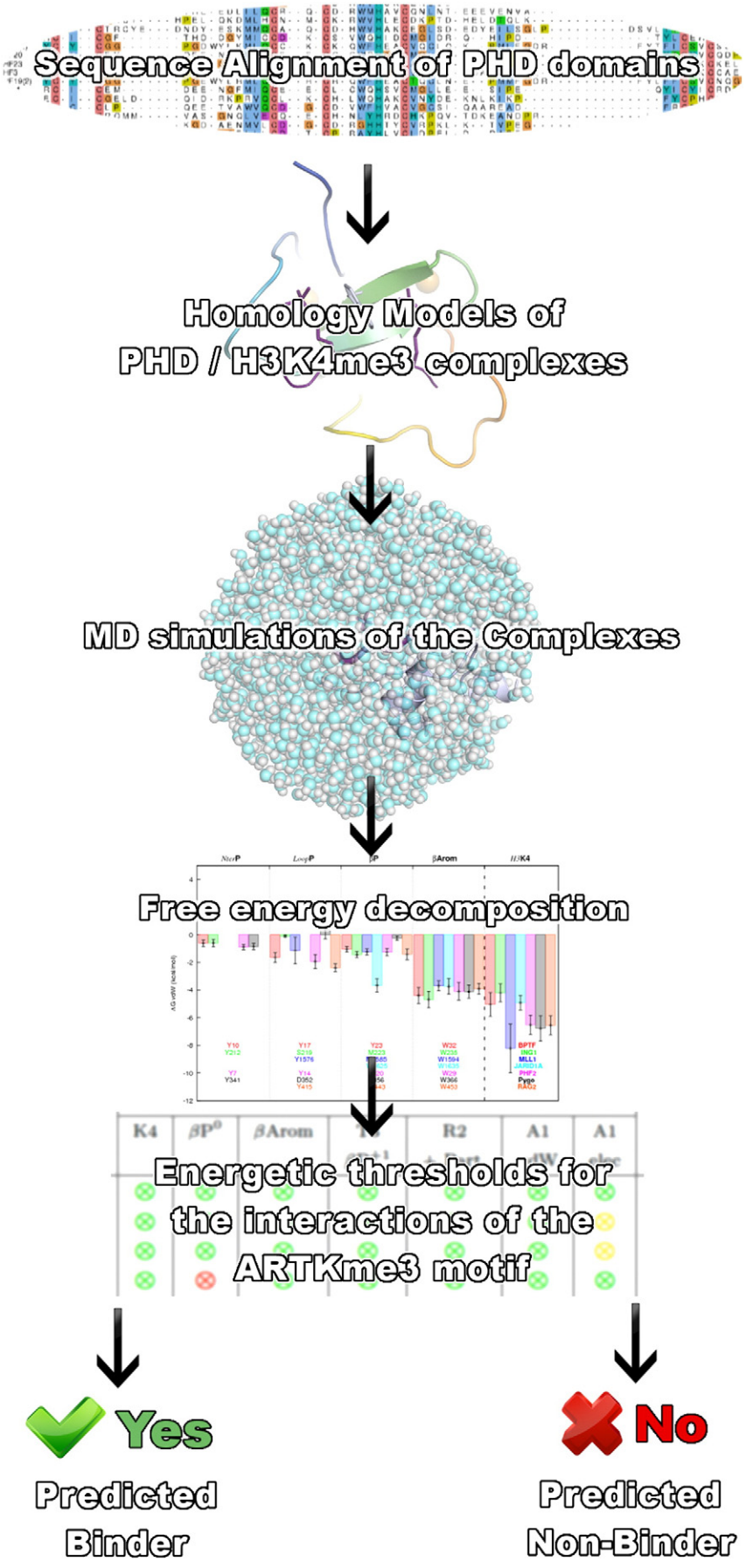


Fig. 2. Summary of the protocol used for the prediction of PHD domains that bind methylated *H3*K4. See Methods section for details.

conserved (see Results section) are not easily identified from sequence comparison alone. Therefore, it was important to perform a 3D homology modeling of the PHDs that possess the aromatic residue βArom in order to assess whether they would likely be specific for *H3*K4. For the homology modeling, we decided to take the following crystallographic structures as templates: *2F6J* (BPTF), *2QIC* (ING1), *2VPE* (Pygopus) and *3GL6* (JARID1A), as their PHDs cover most of the sequence variability of the protein/peptide interface. In the course of procedure validation, when we reconstructed an experimental structure that, in any other case would serve as a template, we removed the corresponding crystallographic structure from the template list. The PHD, the two zinc atoms and the six N-terminal residues of the histone peptide served as the template. The *automodel* module of the Modeller 8v3 program was used to generate models that were then evaluated using the DOPE score. In this study, 50 initial models were generated and the 10 with the best DOPE score were then prepared for molecular dynamics simulation using the protocol described above. For these particular systems, this was sufficient for obtaining conclusive results. However, one is not limited in the number of generated models, which can be adapted to a particular study.

2.3.2. Energetic analysis of modeled PHD domains

In order to rapidly evaluate the binding energies of the complexes, the MD simulations of the model domains were carried out using stochastic boundary conditions. A 25 Å radius sphere of TIP3P water molecules was centered on the C_α of *H3*K4me. The systems were submitted to a 2 ps heating phase at 298 K and a 10 ps equilibration phase using harmonic restraints. The restraints aimed at stabilizing the peptide at the surface of the domain and to reinforce the known interactions that include, *i*) main chain interactions of *H3*K4 and *H3*R2, *ii*) a hydrogen bond between *H3*A1 α-ammonium and the closest carbonyl of L2 loop, and *iii*) a hydrogen bond between *H3*R2 guanidium and the closest acidic residue at positions βP^{+4}, $\beta Arom^{-2}$, and $\beta Arom^{-1}$. Restraints were removed after equilibration and the system was subjected to a 100 ps production phase simulation. With 10 models per domain studied, we thus obtain a cumulative 1 ns of sampling on which the energy analysis was performed. During the molecular dynamics simulations, conformations were stored every 0.5 ps. From this ensemble of structures, we extracted 10 representative conformations for the energetic analysis using the above-described protocol. Based on the results of the study of experimental structures (see Results section), a reference set of interactions (see Fig. 1 for labeling of amino acids) was defined and comprised of, *i*) the binding energies of the methylated lysine and its two main partners (βArom, βP^0), *ii*) the cumulative binding energies of *H3*T3 with βP^{+1}, *iii*) the sum of *H3*R2 binding energies with its interacting partners, *iv*) the sum of electrostatic energies of *H3*A1 α-ammonium and interacting residues of L2 loop, and *v*) the van der Waals term of *H3*A1 side chain. Finally, we ranked the modeled complexes according to the magnitude of their interactions with each amino acid of the methylated H3 tail. For each complex, rather than summing up the contributions to obtain a global score, we considered that all the interactions had to be present simultaneously for the complex to form. We therefore ranked the magnitude of each individual interaction in order to make a prediction as to whether the complex would form or not. If the magnitude of the interaction was (in absolute value) within two thirds of what was observed for known complexes (Table 2), the interaction condition was considered satisfied.

3. Results

3.1. Energetic data on experimental PHD-histone peptide complexes

We carried out an energetic analysis of the methylated histone peptides/PHD complexes listed in Table 1 following the methodology highlighted in the Methods section. The goal of the analysis was to identify recurring interactions across an ensemble of structures, to highlight the physico-chemical basis for the recognition of the peptides by the PHD and to identify the amino acids of the PHD that play a dominant role in the binding of histone peptides. In order to label recurrent interactions in a consistent manner in different PHDs, we developed a name convention based on the structure rather than on the sequence of the different domains, see Fig. 1B for the structure convention and Fig. 1C for the associated sequence alignment. To introduce this convention, we first recognize that in all PHDs, the histone peptide binds as a complementary anti-parallel β-sheet and interacts with several residues of the corresponding β_1 strand of the PHD. The predominant interactions with the PHD are via the first four amino acids of the N-terminal H3 sequence, ARTK. In our convention, we specifically label amino acids of the PHD that interact with these H3 residues.

Amino acids of the PHD β_1 strand will be referred to as βP^x where the *x* makes reference to their position. More specifically, the residue of the PHD that is involved in (two) main chain interactions with the methylated *H3*K4 of the histone peptide is labeled βP^0 and residues immediately C-ter of βP^0 are labeled βP^{+1} till βP^{+4} (cf. Fig. 1C). The residue βP^{+1} is always a hydrophobic residue, with a preference for Ile (see Fig. 1C) while the residue βP^{+2} forms a conserved main-chain interaction with *H3*R2.

Another important residue of the *H3*K4me binding pocket is a conserved aromatic residue lying on strand β_2. It is a Trp in all systems studied and we label it βArom. Other residues of the β_2 strand that mediate side-chain interactions are labeled with respect to this conserved aromatic residue, in particular residues $\beta Arom^{-1}$ and $\beta Arom^{-2}$ that also belong to *H3*R2 pocket. Finally, the binding pocket for the methylated H3K4 side-chain contains residues situated in the N-terminal part of the PHD domain (Nter-P) and in the loop between the β_2 and β_3 strands (LoopP). The buried side-chain of *H3*A1 faces a conserved aromatic residue situated in the C-terminal loop of the PHD domain (Aromα). This name convention allows structurally equivalent amino acids in the different PHDs to have a common label. The positions of these amino acids in the sequence alignment of the different PHDs are shown in Fig. 1C. In the sections below, we describe the energetic data obtained for the different complexes. We focus in particular on the first four amino acids of the H3 histone tail that interact with the PHDs in all experimental complexes and we describe their relevant interactions as identified by the energetic analysis.

Table 2
*H3*A1 and *H3*T3 interactions. Binding energies of *H3*A1 and *H3*T3 and their interacting partners (residues of the L2 loop and βP^{+1}, respectively) in the experimental systems. The contributions of these residues are provided at the backbone/side chain detail and an emphasis is made on the dominant contribution, i.e. electrostatic for *H3*A1, van der Waals for *H3*T3.

		BPTF	ING1	MLL1	JARID1A	PHF2	Pygo	RAG2
*H3*A1	BB Elec.	+3.5	+0.4	+11.7	+0.2	+7.9	+5.1	+11.2
L2-loop	BB Elec.	−3.0 (A35)	−1.0 (G249)	−2.1 (P1614)	−3.2 (A1648)	−2.2 (I45)	−0.6 (A386)	−3.1 (L469)
		−3.6 (I38)		−3.5 (V1617)	−3.7 (E1651)	−1.5 (D46)	−1.7 (A388)	
*H3*A1 + L2-loop	BB Elec.	−3.1	−0.6	+6.1	−6.7	+4.2	+2.8	+8.1
*H3*T3	BB + SC vdW	−3.5	−3.3	−3.7	−3.4	−3.3	−4.0	−3.5
βP^{+1}	BB + SC vdW	−1.8	−1.9	−1.4	−1.6	−2.0	−2.0	−1.9
*H3*T3 + βP^{+1}	BB + SC vdW	−5.3	−5.2	−5.1	−5.0	−5.3	−6.0	−5.4

3.1.1. Recognition of methylated lysine H3K4

The modified *H3*K4 side chain inserts in a pocket formed by residues NterP, LoopP, βP^0 and βArom. An image showing the structure of the binding site for each complex is provided as Supplementary Material (Figure S1). With the exception of *Pygopus*, where the lysine is di-methylated [51], *H3*K4 is tri-methylated in all complexes. In accord with the crucial role of lysine methylation for the formation of these complexes, the energetic analysis (see Fig. 3) indicates that this modified lysine always makes strong contributions to the formation of the complex with binding energies ranging between −5 and −9 kcal/mol. Though part of this contribution relies on backbone interactions (−1.5 to −2 kcal/mol), most of it is due to side chain contacts (−3 to −7 kcal/mol) (Figure S3). Among the residues forming the methylated lysine cleft, the positions βP^0 and βArom usually arise as main interacting partners. The conserved Trp (βArom) interacts with the methylated lysine solely by its side-chain, where it has on average −5 kcal/mol binding energy, making it the strongest contributor of the *H3*K4 pocket (see Fig. 3). Important interactions are also mediated by the PHD residue βP^0. Although this latter residue shows sequence variability (Tyr, Trp, Met or Ala are found at this position), it always makes a significant contribution in terms of binding energy (see Fig. 3) and these contributions originate from both the main chain and side chain contacts. Contributions to the binding energy of the two remaining residues of the methylated *H3*K4 binding pocket (NterP and LoopP) show higher variability. When they are both aromatic, as in the case of BPTF and PHF2, they form an aromatic cage together with the conserved Trp (βArom) around the methylated lysine. However, their energetic contribution remains smaller than that of the conserved Trp. In agreement with their lesser energetic importance, the residues NterP and LoopP are not systematically present in the PHDs studied; for example, they are not present in JARID1A. In two PHDs, RAG2 and Pygo, the energetic contribution of residue LoopP is more important. In RAG2, the LoopP residue forms a somewhat different pattern of interactions that involve not only the methylated *H3*K4, but also the peptide residue *H3*Q5, which contributes to its conformational stabilization and to more favorable interactions. RAG2 is somewhat atypical for a PHD as it possesses an N-terminal sequence insertion and its zinc ions are coordinated by a Cys3-His2-Cys2-His motif (see Fig. 1c). Another atypical interaction of residue LoopP is observed in Pygo, the sole PHD studied that binds di-methylated *H3*K4. Pygo harbors an Asp at the LoopP position that forms a salt bridge with *H3*K4me2 ζ-ammonium and thus ensuring selectivity toward the Kme2 versus Kme3.

3.1.2. Optimized van der Waals interactions of H3T3

The *H3*T3 residue has a conserved binding mode in PHDs, with its hydroxyl group pointing toward the solvent and the γ-methyl group facing the residue βP^{+1}, which is always a hydrophobic residue and by preference an Ile (see Fig. 1C). The PHD βP^{+1} residue mediates significant van der Waals interactions with the γ-methyl group of *H3*T3 and the cumulative van der Waals contribution of *H3*T3/βP^{+1} is highly similar in all complexes, ranging between −5 and −6 kcal/mol (See Fig. 1C). The side chain of *H3*T3, on the other hand, seldom forms a hydrogen bond with the PHD, with the exceptions of Pygo and MLL1. The polar residue interacting with *H3*T3 in these latter examples is located in helix α_2, but despite its rather strong contribution in both cases, polar interactions of *H3*T3 do not appear as a recurring pattern.

3.1.3. Variable but favorable binding modes for H3R2

*H3*R2 is anchored in a cleft of the PHD surface that runs parallel to the *H3*K4 cleft. The separation between the two clefts is formed by the conserved Trp residue (βArom). *H3*R2 makes conserved main chain hydrogen bonds with the backbone of residue βP^{+2} and the two side chains are always adjacent. The binding energies of these two residues are favorable in all cases and, as can be seen in Fig. 4, that of βP^{+2} is rather small (≈−2 kcal/mol on average). The strongest interactions of *H3*R2 (ranging between −2 and −6 kcal/mol) are made with the side chains of residues that can be located at four positions: βP^{+4} $\beta Arom^{-2}$, $\beta Arom^{-1}$ and also sometimes at βP^{+2}. The residue βP^{+2} has a significant impact on the size of *H3*R2 cleft and thus, with which residues the arginine can interact. In a few systems (BPTF, ING1, TAF3 and Yng1), βP^{+2} is either a Gly or an Ala. This results in a rather large cleft where the side chain of *H3*R2 is fully accommodated and can easily form strong electrostatic interactions with acidic residue(s). In this configuration, the conserved Trp is sandwiched between the side chains of *H3*K4me and *H3*R2 (see Fig. 1b). The residue βP^{+2} can, in some proteins, be a bulky hydrophobic such as a Leu (Pygo) or an aromatic like a Tyr (RAG2). In these cases, the *H3*R2 cleft is nearly nonexistent (Figure S2c and S2d of Supplementary Material). In this latter case, only the aliphatic carbons of *H3*R2 side chain contact the PHD, while the charged guanidium group points toward the solvent. The

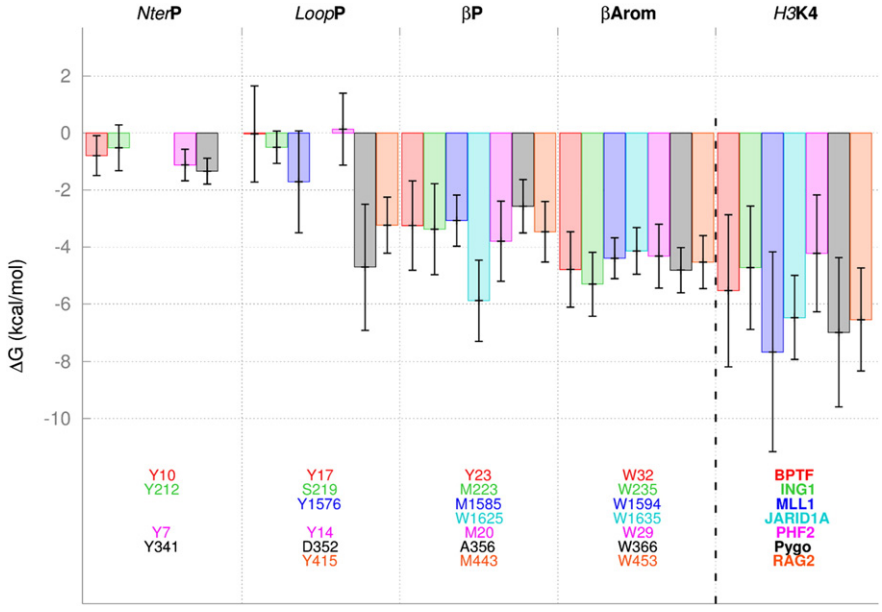


Fig. 3. Binding energies (kcal/mol): *H3*K4 pocket. Total binding energies of the residues forming the methylated lysine pocket (*Nter*P, *Loop*P, βP^0, βArom, and *H3*K4me) for the seven representative systems.

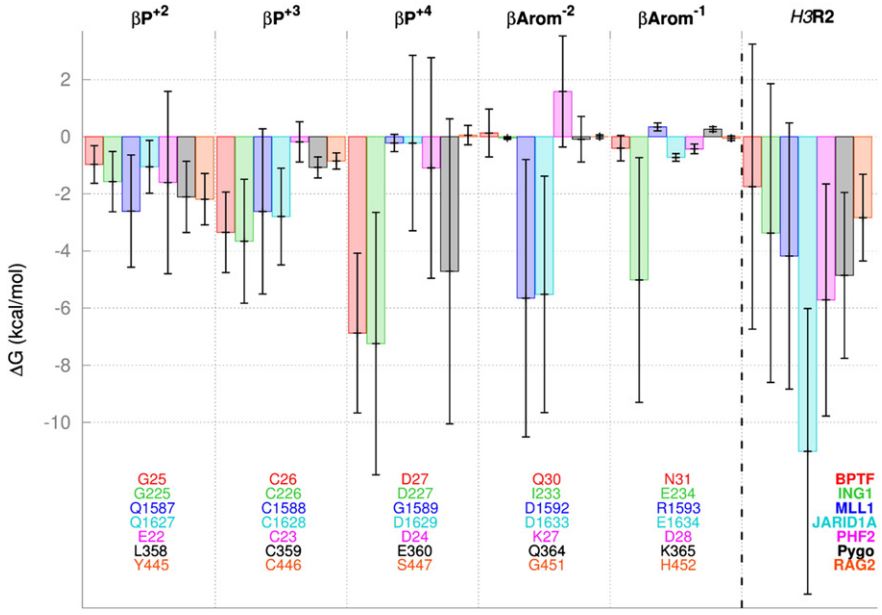


Fig. 4. Binding energies (kcal/mol): *H3*R2 cleft. Total binding energies of the residues forming *H3*R2 pocket (βP^{+2}, βP^{+3}, βP^{+4}, $\beta Arom^{-2}$, $\beta Arom^{-1}$ and *H3*R2) for the seven representative systems.

observation that arginine methylation has no effect on binding affinity is coherent with purely hydrophobic interactions for *H3*R2 in these two systems [51,52]. An intermediate situation is observed when βP^{+2} is either a Gln or a Glu. In JARID1A and MLL1, the βP^{+2} Gln allows the existence of a small cleft and in both cases, the guanidium forms a strong electrostatic interaction with an Asp at $\beta Arom^{-2}$ (Fig. 4, Figure S2c and Figure S2e). In PHF2, as well as in PHF8, *H3*R2 interacts mostly with the Glu in βP^{+2}. Present in many of the crystallographic structures is a bridging water molecule that provides a hydrogen bond between the main chains of *H3*R2 and the conserved zinc-binding Cys at βP^{+3}. In the three systems where this water bridge is stable during trajectories (BPTF, ING1 and JARID1A), the βP^{+3} Cys has a favorable binding energy. PHDs can thus use a strikingly high number of ways to accommodate the *H3*R2 side chain, but it generally involves at least one strong electrostatic interaction with the guanidium. Domains binding *H3*R2 only through van der Waals interactions (Pygo and RAG2) can thus be seen as particular cases that have evolved in order to be able to recognize H3 tails methylated on both *H3*R2 and *H3*K4.

3.1.4. Specific anchoring of the N-ter H3A1

An important aspect of the molecular recognition of H3 tail peptides by PHDs is that, besides *H3*K4 and *H3*R2, they interact with another positively charged group: the N-terminal extremity of *H3*A1. The α-ammonium group of *H3*A1 is most often hydrogen bonding to carbonyl groups of amino acids of the L2 loop. Besides these recurring interactions, the *H3*A1 α-ammonium also forms H-bonds with an acidic residue in the βP^{+4} position. It is interesting to note that the acidic amino acid at the conserved βP^{+4} position is also involved in stabilizing the guanidium of *H3*R2, and thus appears as an important stabilizing interaction for the recognition of H3 tail. However, this interaction is not systematically present in all the PHDs studied. For example, residue βP^{+4} is a Gly and a Tyr in MLL1 and RAG2, respectively. In the case of MLL1, compensating interactions with an acidic residue at another nearby position (Glu1615 on L2-loop) are found, but not in RAG2. This lack of stabilization might explain the lower affinity of RAG2 for H3 tail as compared to the other systems studied (see Table 1). As a result of these diverse interaction patterns, the electrostatic contribution from the charged N-terminal is rather variable. Besides the charged N-terminal, the methyl group of *H3*A1 forms recurring van der Waals interactions with a well conserved aromatic residue that we label $Arom_{\alpha}$ (see Fig. 1a). Overall, the methyl groups of *H3*A1 and *H3*T3 appear essential for the hydrophobic anchoring of the peptide given that, in the systems studied, they respectively mediate on average 4 and 4.5 hydrophobic contacts with a significant stabilizing interaction.

3.1.5. Residues C-ter of methylated H3K4 mediate few interactions with PHD domains

While we identified conserved partners for the four N-terminal residues of the *H3*K4 peptide, the interactions of the C-terminal residues are more variable. Of these residues, *H3*Q5 is the only one having a relatively well-conserved energetic profile that shows a significant van der Waals contribution (see Table S1). This interaction occurs chiefly with the residue located at position βP^{-1}, where different types of amino acids (Phe, Pro, Ala, Asp/Glu) can be found (see Fig. 1B for the position of this residue in the structure of the domain). The residues situated further toward the C-terminal of the histone peptide do not display any conserved interactions. In most cases, only six or seven residues of the peptide mediate interactions with the PHD. Exceptions are found for ING4 and ING5, in which more C-terminal peptide residues are apparent in the crystal structures. In these two systems, *H3*R8 makes a stable salt bridge with an acidic residue at the extreme N-terminal end of the PHD ($NterP^{-3}$). Despite the fact that this acidic residue is well conserved among the systems studied, its interaction with *H3*R8 is not frequently observed and is probably not important for the recognition of *H3*K4 site. From the energetic analysis, the PHDs studied recognize the N-terminal of H3, methylated on *H3*K4, as a linear motif [64], and the combination of the first four amino acids of the sequence, rather than merely the methylation mark on *H3*K4, determines the selectivity and pattern of recognition. Fig. 2 intends to summarize the structural analysis we detailed over the last few paragraphs.

3.2. Position-specific energy thresholds and validation of the prediction protocol

Although a significant amount of structural data is available for complexes of PHDs with an *H3*K4 peptide, there remain large numbers of PHD-containing proteins for which no structural data are available. Some of these proteins contain multiple PHDs, as in the Histone Methyl Transferase enzyme MLL3, which contains 6 such domains. Significant efforts are being made to identify new modified histone tail binders,

thus making the prediction of all PHDs that could recognize the methylated *H3*K4 site of particular interest. Recently, an experimental proteome-wide analysis of *S. cerevisiae* identified several PHD fingers that bind histone H3 methylated on *H3*K4 [16]. However, there is no corresponding data for human PHDs. Our interest is to assess whether the energetic analysis performed on experimental structures of PHD/histone peptide complexes can be used to identify the histone binding propensities of yet uncharacterized PHDs. For that purpose, we set-up a protocol based on the construction of homology models of the complexes, followed by an energetic analysis, as described in the Methods section. In this section, we assess the validity of the protocol by modeling point mutations into PHDs that are known to affect *H3*K4me binding and predict the resulting effect. We also analyze the *S. cerevisiae* genome in order to predict the *H3*K4me binding propensity of *S. cerevisiae* PHD fingers and subsequently compare our predictions to the experimental data.

3.2.1. Comparison of interactions on experimental and modeled structures of PHDs

The first step of the computational proteomics protocol was to build homology models of the PHD domains for which only the sequence is available. In order to assess the influence of the structural uncertainty introduced by the modeling procedure, and to evaluate its impact on the predicted energetics of recognition, we constructed homology models for a subset of structurally known PHDs (BPTF, ING1, JARID1A, PHF8, PYGO2, TAF3 and YNG1) after having removed the target structure from the list of templates. From these models, the energetic contributions of all amino acids were computed and compared to the results obtained from the analysis of the experimental complexes (see Table 3). The correlation between the energies obtained on the experimental and modeled complexes was good with a correlation coefficient of 0.945. In particular, average values of the energies and standard deviations are in very good agreement (see Table 3). Overall, the quality of the results validates the computational procedure for constructing models of the various complexes.

3.2.2. Prediction of the effect of point mutations on the affinity of PHD domains for methylated H3 peptides

In several of the structural studies of PHD-*H3* peptide complexes, point mutations in the PHDs were introduced in order to confirm the importance of the intermolecular interactions visible in the structure. We constructed model structures of some point mutants using the same modeling procedure and we evaluated the impact of the mutation on the energetics of recognition (see Table 4). In Pygo, the mutations A → E of the residue βP^0 destabilized the complex about 100 fold, in agreement with the unfavorable contribution of this residue in the energetic analysis. Likewise, mutation βP^{+1} I → R shows no detectable binding, which results from a lack of anchoring of the *H3*T3 residue, and mutation βP^{+4} E → A, which also results in no binding, is predicted to destabilize the anchoring of *H3*R2. In ING2 and TAF3, mutation of the acid residue in position βP^{+4} destabilizes the complex, in agreement with our models that predict that it disrupts the electrostatic anchoring of the N-terminal residue. The importance of the aromatic residue (Trp), which separates the methylated-K4 and R2 pockets (labeled βArom), is assessed by its sequence conservation and apparent from the structural information. This importance is further shown by the energetic analysis presented above, which highlights its significant contribution. Mutation of this residue to a non-aromatic residue has been performed, for example in Pygo (W → E, [51]), leading to loss of binding of the methylated peptide. More conservative mutations, namely Trp to Phe or Tyr, have been reported with different outcomes: no effect on binding of methylated *H3*K4 for TAF3, (W → Y) and about 10 fold loss in affinity in BPTF (W → F). It must be noted that very few natural sequences of PHD fingers have an aromatic residue other that a Trp at that position (one example: Pygo of *Drosophila* in which it is a Phe [51]).

3.2.3. Analysis of S. cerevisiae PHD domains

A proteome wide analysis for PHDs that bind methylated histones was performed experimentally for *S. cerevisiae* [16]. A total of 14 PHD finger-containing proteins were tested and 8 domains were found to bind methylated *H3*K4. To further assess the predictive ability of our modeling protocol, we constructed models and assessed the methylated histone binding potential of several *S. cerevisiae* PHD domains that were previously studied experimentally. The computational predictions, together with the experimental data, are summarized in Table 5. All the energetic interactions of the PHD domains BYE1, SPP1, ING1, SET3, PHO23, CTI6 with the methylated H3 tails are similar to what is observed in known binders (see Table 5). The prediction that they bind the methylated histone tail is in agreement with the experimental data. SET4 shows a large deviation in the energetics of *H3*R2 binding. This is linked to the presence of an Asn at position βP^{+4} instead of an acid, which would better anchor the Arg side chain. It is therefore predicted as being a nonbinder, in agreement with the experimental data. The other *S. cerevisiae* PHD domains experimentally found to be non-binders [16] do not have an aromatic residue at the position βArom. They were therefore correctly predicted as non-binders at the bioinformatics filter stage (see Fig. 5) and are not displayed in Table 5. The only divergence between the experimental and computational

Table 3
Reference binding energies of the main hotspots obtained with the prediction protocol (top), and from the study of the crystallographic structures (bottom). All energies are given in kcal/mol.

		Values from prediction protocol								
Residue(s)		Energy	BPTF	ING1	JAD1A	PHF8	Pygo2	TAF3	Yng1	Average
*H3*K4	BB + SC	Tot.	−5.3	−4.5	−5.4	−5.5	−7.5	−4.9	−3.7	−5.3 ± 1.2
βP^0	BB + SC	Tot.	−3.7	−3.2	−3.9	−3.1	−2.1	−3.3	−3.2	−3.2 ± 0.6
βArom	SC	Tot.	−5.1	−5.2	−4.3	−4.1	−4.7	−4.9	−4.9	−4.7 ± 0.3
*H3*T3 + βP^{+1}	BB + SC	Tot.	−4.8	−5.4	−5.0	−5.0	−4.9	−5.0	−4.6	−5.0 ± 0.2
*H3*R2 + partners	BB + SC	Tot.	−6.6	−15.2	−16.6	−16.8	−12.0	−10.4	−12.0	−12.8 ± 3.3
*H3*A1	SC	vdW	−1.5	−1.5	−1.7	−1.4	−1.7	−1.5	−1.3	−1.5 ± 0.1
*H3*A1 + partners	BB	Elec.	−0.7	−0.2	−3.7	+0.4	−1.6	−1.4	+3.4	−0.5 ± 2.1
		Values from crystallographic structures								
Residue(s)		Energy	BPTF *2F6J*	ING1 *2QIC*	JAD1A *3LG6*	PHF8 *3KV4*	Pygo2 *2VPE*	TAF3 *2K17*	Yng1 *2JMJ*	Average
*H3*K4	BB + SC	Tot.	−5.5	−4.7	−6.5	−3.4	−7.0	−5.7	−5.1	−5.4 ± 1.0
βP^0	BB + SC	Tot.	−3.2	−3.4	−5.9	−1.3	−2.6	−4.2	−4.1	−3.5 ± 1.3
βArom	SC	Tot.	−4.8	−5.3	−4.1	−4.7	−4.8	−5.0	−4.5	−4.7 ± 0.3
*H3*T3 + βP^{+1}	BB + SC	Tot.	−5.3	−5.4	−4.6	−3.3	−6.7	−4.5	−4.7	−4.9 ± 0.9
*H3*R2 + partners	BB + SC	Tot.	−9.4	−18.4	−19.8	−13.3	−11.9	−12.9	−11.1	−13.8 ± 3.5
*H3*A1	SC	vdW	−1.4	−1.6	−1.4	−1.5	−1.3	−1.4	−1.2	−1.4 ± 0.1
*H3*A1 + partners	BB	Elec.	−3.1	−0.6	−6.7	+0.8	+2.8	−1.3	+3.5	−0.6 ± 3.2

Table 4
Relative deviations of the main energies from the average values of Table 3 for point mutations on the experimental systems. Energies lower than two thirds of the reference are represented in green, between one and two thirds in yellow, between zero and one third in orange, and positive values are shown in red. In the particular case of *H3*A1 electrostatic interactions, as the reference is close to zero, we use a 3 kcal/mol increment between each color. βPHD stands for the residues βP^0 to βP^{+4}, plus residues $\beta Arom^{-1}$ and $\beta Arom^{-2}$, that are the main partners of the four N-Ter residues of histone H3.

Prot. Name	Point Mutation	βPHB Sequence	Deviation from average (kcal/mol)							Affinity reduction	Ref.
			K4	βP^0	βArom	T3 + βP^{+1}	R2 + Part.	A1 vdW	A1 Elec.		
BPTF	βArom(W→F)	YIGCD/QN	✔	✔	✔	✔	✔	✔	✔	20-fold	[24]
TAF3	βP^{+4}(D→A)	MIGCA/DD	✔	✔	✔	✔	✔	✔	✖	30-fold	[53]
ING2	βP^{+4}(D→A)	MIGCA/IE	✔	✔	✔	✔	✔	✔	✖	35-fold	[25]
Pygo	βP^0(A→E)	EILCE/QK	✔	✖	✔	✔	✔	✔	✔	100-fold	[51]
Pygo	βP^{+1}(I→R)	ARLCE/QK	✔	✔	✔	✖	✔	✔	✔	No binding	[51]
Pygo	βP^{+4}(E→A)	AILCA/QK	✔	✔	✔	✔	✖	✔	✔	No binding	[51]

predictions is for ECM5, for which an interaction was not detected in the experimental assays, but which presents an energetic profile similar to that of known binders. The explanation put forward in the experimental paper [16] for the lack of binding to ECM5 is that of steric hindrance, which is not coherent with our model of the structure in which no such hindrance is observed. This would suggest that further investigation is warranted for this system.

3.3. Identification of human PHD domains that bind methylated H3K4 peptides

From the modeled structures of human PHDs and the subsequent energetic analysis (see Methods), binding propensities for methylated *H3*K4 peptides were predicted. As was apparent from the *S. cerevisiae* data, the conserved tryptophan residue βArom is required for the recognition of the methylated lysine side-chain. Therefore, from the list of all human PHDs (PFAM/PF00628 data base) we filtered out all those PHDs that do not possess a Trp at the βArom position (see Methods). This left 22 sequences of human PHDs that are potential binders of the methylated *H3*K4me tail. Of those, we identified ten PHDs as having energetic interactions with the methylated histone that are similar to those of known PHDs with micromolar affinity for the methylated *H3*K4 peptide. These are therefore predicted interactors (see Table 6). We then identified eight PHDs for which at least one of the important interactions differs from the known binders (Table 6) and four that have a sequence insertion in the L2 loop, which makes the structural predictions less reliable.

3.3.1. Human PHD domains predicted as H3K4me binders

Our computational protocol identified several PHDs that were annotated as *H3*K4me binders, but without any explicit experimental support, as well as PHDs for which experimental information became available during the course of this study. Here we run through the list of proteins that we identified as *H3*K4me binders. ING3 is a component of NuA4 histone acetyl-transferase (HAT) complex, which is involved in transcriptional activation and epigenetic modifications [66]. The solution structure of ING3 PHD in *apo* form is available (PDB *1X4I*). The ING3 PHD displays a significant (92%) sequence homology to the PHD domain of the related protein ING2, which is known to bind tri-methylated *H3*K4 [26]. Based on this, it was therefore inferred that ING3 would act as a methylated *H3*K4 binder [66]. The results from the protocol described here firmly support this prior hypothesis. KDM7 (also know as KIAA1718 and JHDM1D) (PDB: *3KV5*, *3KV6*) belongs to the family of Jumonji histone demethylases [67] and it contains a PHD that is homologous to the PHD of PHF8 (same βPHD sequences, see Table 6 but $\beta Arom^{-2}$ is a Lys instead of a Gln). KDM7 was shown to bind tri-methylated *H3*K4, but only its apo structure was available [50]. From our data, we propose a 3D structure for the complex that is coherent

Table 5
Relative deviations of the main energies from the average values of Table 3 for PHD domains of *S. cerevisiae*. Energies lower than two thirds of the reference are represented in green, between one and two thirds in yellow, between zero and one third in orange, and positive values are shown in red. In the particular case of *H3*A1 electrostatic interactions, as the reference is close to zero, we use a 3 kcal/mol increment between each color. βPHD sequence stands for the residues βP^0 to βP^{+4}, plus $\beta Arom^{-1}$ and $\beta Arom^{-2}$, which are the main partners of the four N-Ter residues of histone H3.

Uniprot Code	Protein			βPHD Sequence	Deviation from average (kcal/mol)						
	Name	Residues	Organism		*H3*K4	βP^0	βArom	*H3*T3 + βP^{+1}	*H3R2* + *Part.*	*H3*A1 + Part.	*H3*A1 Elec.
P36106	BYE1	74–134	*Yeast*	MVQCD/DT	✔	✔	✔	✔	✔	✔	✔
Q03012	SPP1	24–72	*Yeast*	MVGCD/DD	✔	✔	✔	✔	✔	✔	✔
P38806	Yng2	224–271	*Yeast*	MVACD/YE	✔	✔	✔	✔	✔	✔	✔
P36124	SET3	119–166	*Yeast*	TIQCD/NR	✔	✔	✔	✔	✔	✔	✔
P50947	PHO23	282–329	*Yeast*	MVGCD/LE	✔	✔	✔	✔	✔	✔	✔
Q03214	ECM5	1240–1290	*Yeast*	MVECE/KE	✔	✔	✔	✔	✔	✔	✔
Q08923	CTI6	74–123	*Yeast*	FIQCE/SS	✔	✔	✔	✔	✔	✔	✔
P42948	SET4	162–210	*Yeast*	FIQCN/KT	✔	✔	✔	✔	✖	✔	✖
Q9V9W8	Pygo2	749–805	*Drosi*	AVFCE/NF	✔	✔	✖	✔	✔	✔	✔

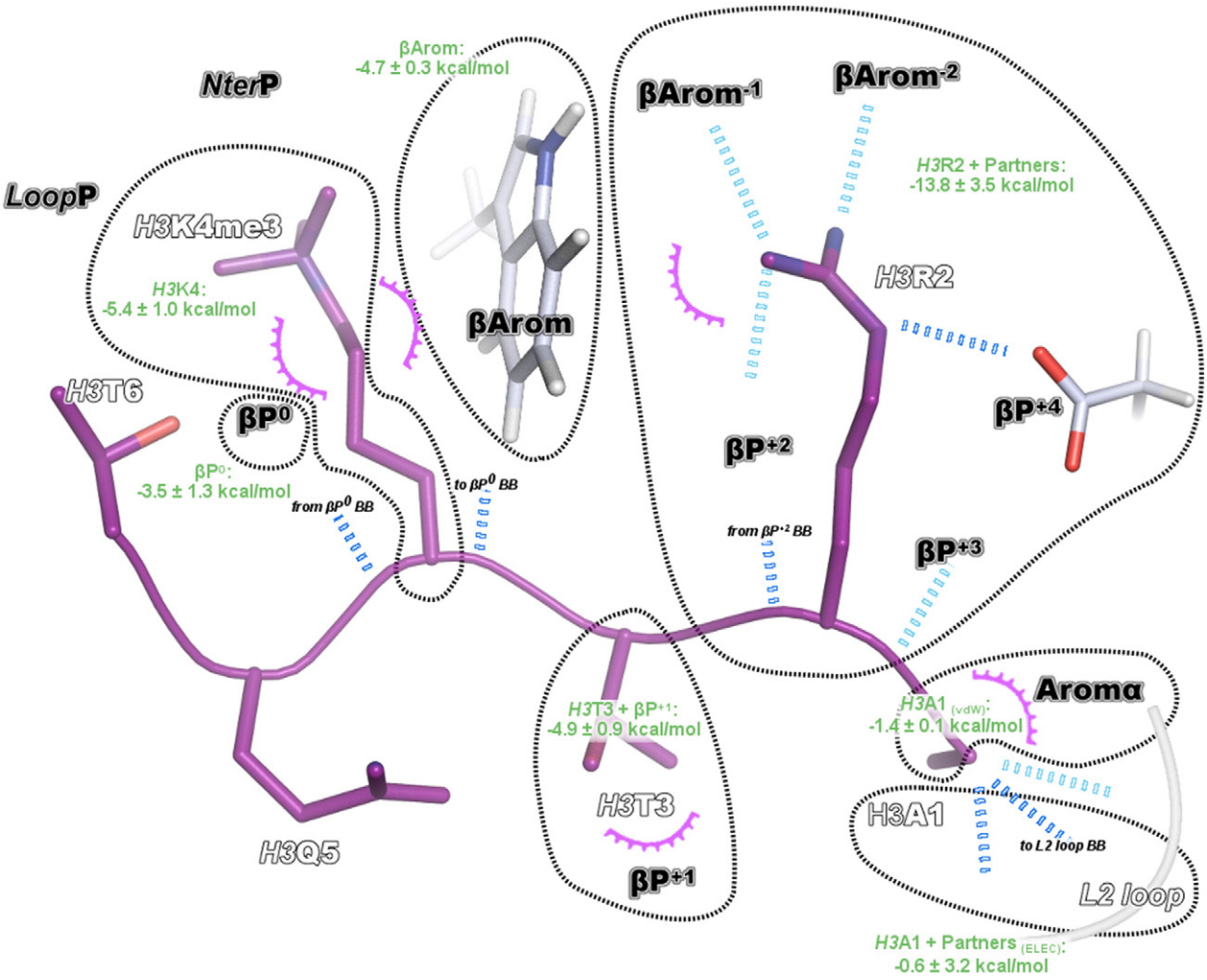


Fig. 5. Summary of the interactions between *H3*K4 peptide and the PHD domain in the experimental structures. The histone peptide is shown in violet (backbone atoms are not displayed). The most common hydrogen bonds are shown with dark blue dots, while those observed more occasionally are represented using light blue dots. The magenta semi-circulars represent residues involved in van der Waals interactions. Finally, the residues or groups of residues mediating the most stable interactions are circled with black dots, and the energies indicated in green are those of Table 3.

with all available biochemical and biophysical data and our analysis confirms the binding of KDM7 PHD domain to tri-methylated histone *H3*K4. DIO-1 (or DIDO-1, Death Inducer Obliterator) is annotated as a putative transcription factor [66,68]. The Dido gene encodes three protein isoforms generated through alternative splicing and all three contain the PHD domain. Several recent studies point out the functional role of DIO-1 in the maintenance of stem cell integrity, as well as its implication in different cancers (see for example [69,70] and references cited therein). Predictions from the protocol presented here are in full agreement with recent functional and structural data on the interactions of DIO-1 with a methylated *H3*K4 peptide [69,71]. Besides the above predictions that have been independently confirmed, we also identified in the absence of structural information *H3*K4 binding PHDs in proteins that have been associated with epigenetic modifications. For example, CpG-binding protein is a transcription factor that exhibits a binding specificity for CpG unmethylated motifs. It is a component of the mammalian Set1 histone H3-Lys4 methyl-transferase complex. PHF20 (PHD finger protein 20) is part of the "male absent on the first" (MOF) lysine acetyl transferase complex, which acetylates histone H4 and also p53 (see [72] and references cited therein). ASH1-like protein (histone lysine N methyl transferase ASH1L) has been shown to specifically methylate *H3*K36 [73]. At this stage, the functional characterization is still rather poor for other predicted *H3*K4 binders, but our results provide leads for further investigations. TCF19 (Transcription Factor 19, previously known as SC1) is annotated in Uniprot as a potential trans-activating factor that could play an important role in the transcription of genes required for the later stages of cell cycle progression [66,74]. It is located in the HLA locus on chromosome 6 and, so far, has not been documented as binding methylated histones *H3*K4. Overall, its function is poorly characterized at the molecular level, however, genetic studies have suggested associations with type-I diabetes [75] and chronic hepatitis B [76]. Interestingly, the second susceptibility locus identified in the latter hepatitis study corresponds to a histone methyltransferase gene (euchromatic histone-lysine-methyltransferase 2 EHMT2), suggesting possible *H3*K4me binding ability. PHF20L1 (PHD finger protein 20-like protein 1) is a protein related to PHF20 for which little information is available at this stage, but our results suggest that it presents *H3*K4me binding capability. There is, however, one PHD domain that we predicted as a binder for H3K4, but for which information to the contrary recently became available. This PHD domain is that of PHF3. PHF3 (also know as KIAA0244) is ubiquitously expressed in normal tissues including brain. The function of PHF3 is poorly characterized at the molecular level, but its expression has been shown to be altered in glioblastoma [77]. Genetic studies have also suggested a PHF3 association with alcohol dependence [78]. In a recent study [69,71], no binding was detected by NMR between PHF3 PHD domain and a methylated *H3*K4 peptide. The authors suggested that the difference in binding affinity between the closely related DIO (see above) and PHF3 PHDs was

Table 6
Relative deviations of the main energies from the average values of Table 3 for human PHD domains. Energies lower than two thirds of the reference are represented in green, between one and two thirds in yellow, between zero and one third in orange, and positive values are shown in red. In the particular case of *H3*A1 electrostatic interactions, as the reference is close to zero, we use a 3 kcal/mol increment between each color. βPHD sequence stands for the residues βP^0 to βP^{+4}, plus $\beta Arom^{-1}$ and $\beta Arom^{-2}$, which are the main partners of the four N-Ter residues of histone H3.

Uniprot	Protein		βPHD Sequence	Deviation from average (kcal/mol)						
Code	Name	Residues		*H3*K4	βP^0	βArom	*H3*T3 + βP^{+1}	*H3R2 + Part.*	*H3*A1 + Part.	*H3*A1 Elec.
Q9Y242	TCF19	295–342	WVQCD/DV	✔	✔	✔	✔	✔	✔	✔
Q9UGL1	$KDM5B_3$	1492–1538	WVQCD/NQ	✔	✔	✔	✔	✔	✔	✔
Q9NXR8	ING3	362–409	MVGCD/IE	✔	✔	✔	✔	✔	✔	✔
Q6ZMT4	KDM7	39–88	MIECD/KD	✔	✔	✔	✔	✔	✔	✔
Q92576	PHF3	719–772	MVGCG/DD	✔	✔	✔	✔	✔	✔	✔
Q9BTC0	DIO–1	270–322	MICCD/EE	✔	✔	✔	✔	✔	✔	✔
Q9P0U4	CpG–BP	28–76	MIGCD/NE	✔	✔	✔	✔	✔	✔	✔
Q9BVI0	PHF20	654–700	MIQCE/QC	✔	✔	✔	✔	✔	✔	✔
Q9NR48	ASH1–like	2587–2631	MIQCD/MV	✔	✔	✔	✔	✔	✔	✔
Q86U89	PHF20L1	70–116	MIQCE/LC	✔	✔	✔	✔	✔	✔	✔
Q9UMN6	$MLL4_3$	1337–1396	MMQCA/DH	✔	✔	✔	✔	✔	✔	✖
Q9BUL5	PHF23	341–387	MIECS/GT	✔	✔	✔	✔	✔	✔	✖
Q8IZD2	MLL5	120–166	MICCD/SV	✔	✔	✔	✔	✖	✔	✔
Q8NEZ4	$MLL3_6$	1086–1139	ILQCR/DR	✔	✔	✔	✔	✖	✔	✖
Q5T6S3	$PHF19_2$	194–249	MLQCY/RQ	✔	✔	✔	✔	✖	✔	✖
Q8NEZ4	$MLL3_3$	466–520	MLHCN/KR	✖	✔	✔	✔	✖	✔	✖
Q9Y483	$MTF2_2$	203–255	MLQCC/KQ	✔	✔	✔	✔	✖	✔	✖
Q149N8	SHPRH	660–709	RVQCL/HL	✔	✖	✔	✔	✖	✔	✖

linked to the substitution Y/Q in the methylated lysine binding pocket (at position labeled NterP). As discussed in the Methods section, while the position NterP is part of the aromatic cage that surrounds the methylated lysine, it is not the main energetic contributor to its binding. In the different systems studied, the amino acid labeled NterP has highly variable contributions and therefore was not considered in devising the energy threshold for binding. This may be at the origin of the poor prediction for PHF3, although further investigations would be necessary to fully establish the origin of the DIO/PHF3 differences.

Of the predicted binders, only two domains (ING3 of the ING family and $JARID1B_3$) have a high sequence identity (76 and 64%, respectively) with a template structure, while the identity is between 50 and 60% for the others. We saw that residues of the segment βP^0 to βP^{+4} (which can be referred to as the βPHD sequence) mediate extensive interactions with the histone peptide. If we compare the composition of this portion of sequence in our predicted and in the known binders (see Table 5 and Figure S4), we observe many similarities, though residues that do not correspond to the sequence of known binders are also found. At the βP^0 position, the predicted binder domains primarily contain a Met, but Trp is also observed at this position in the known binders. Recall that in the case of *S. cerevisiae*, we had correctly predicted that Tyr or Phe, residues not observed in known structures, would also permit binding. At βP^{+1} position, the main interacting partner of *H3*T3, there is always a hydrophobic residue in the PHD alignment (Ile, Leu, Val, Met). At the position βP^{+2}, known binders have side chains that differ significantly in size and physico-chemical properties (Gly, Ala, Leu, Glu, Gln, Tyr), and as discussed above, this residue has a modulating influence on the number and the type of interactions that *H3*R2 can form with the PHD. In the predicted binders, this residue is frequently a Gln, as in the template JARID1A, but Gly, Ala, Glu, and Cys are also observed. These residues present neither a steric obstacle nor charge repulsion to the insertion of *H3*R2 in the groove on the PHD surface. Finally, in the predicted binders, the guanidium group of *H3*R2 can always form at least one electrostatic interaction with acidic residue(s) at positions βP^{+4}, $\beta Arom^{-1}$ or $\beta Arom^{-2}$. Based on this observation, we would expect the binding affinity of the predicted complexes to be affected by post-translational modifications of the arginine side chain (i.e.: methylation).

3.3.2. Human PHD domains predicted as poor/non binders of H3K4me

As discussed before, the absence of an aromatic residue, mostly a Trp, at the position βArom hampers the binding of the methylated lysine and therefore, PHDs that do not have this conserved residue are predicted to not bind the methylated histone. However, as shown for *S. cerevisiae*, this condition, although necessary, is not sufficient to ensure the binding of the *H3*K4me3 histone tail. We identified a number of PHD domains for which at least one of the other interactions characteristic of complexes with methylated *H3*K4 is not optimally formed and that we would therefore predict as poor or nonbinders for methylated *H3*K4 (it must be kept in mind that the affinity threshold for detection of interactions depends on the experimental method used and therefore

the distinction between poor and non-binders remains somewhat arbitrary).

The position of the poorly formed interaction differs in the different complexes. For the third PHD ($MLL4_3$) of the histone-lysine N-methyltransferase MLL4, as well as for the PHD finger of PHD finger protein 23, binding does not seem optimal because the N-Terminal alanine is predicted to be poorly anchored. The first and second PHDs of MLL4 were not modeled, as they do not have the required sequence characteristic for binding a methylated lysine. For the PHD domain of histone-lysine N-methyltransferase 2E (KMT2E, also known as MLL5), a poor anchoring of the arginine *H3*R2 was observed in the modeled structure. And for several PHDs, more than one interaction point is lost; poor anchoring of both *H3*A1 and *H3*R2 was observed for both MLL3(6), a histone methyltransferase that contains six PHD fingers, and for PHF19(2) and MTF2(2), two polycomb group proteins that each contain two PHD domains. Finally, the unique PHD domain of SHPRH, an E3 protein ubiquitin ligase, looses interactions of *H3*A1, *H3*R2 and *H3*K4me. For this latter case, independent confirmation that the PHD domain does not bind methylated *H3*K4 appeared recently [79].

4. Discussion and conclusion

Post-translational modifications of histones play an essential role in epigenetic modifications and large efforts are currently undertaken to identify the cellular partners of modified histones and to decipher their mode of action. For that purpose, the development and validation of efficient proteomics methods dedicated to the identification of modified histones partners are of paramount importance. In this work, we developed an approach based on a combination of a bioinformatics filter, homology modeling and molecular dynamics simulations that can be used in a rather straightforward way to make binding predictions. For validation, we focused on a family of protein domains for which there is extensive structural information, namely PHD fingers. We showed that methods we developed here can predict their binding propensity toward methylated *H3*K4, a critically important post-translational modification. Currently, several 3D structures of *H3*K4me3 peptides bound to PHD fingers are available (see Table 1). Biophysical characterization of the interaction of the methylated histone peptide with PHDs indicates that they form low affinity complexes (1–10 μM, see Table 1). Molecular determinants for the recognition of methylated lysines over unmodified lysines have been established by structural studies, where it was noticed that an aromatic cage often surrounds methylated lysines [42], while salt-bridges mediated by acidic residues are involved in the recognition of unmodified Lys [27]. However, this observation is not sufficient to identify protein domains that bind methylated Lys from sequence alone, as the spatial arrangement of the aromatics is essential for recognition. Moreover, the problem is more often the identification of methylation marks in a particular sequence context, such as the N-terminal of histone H3 (*H3*K4) studied here. The molecular basis for the sequence specific recognition has not been as thoroughly investigated as that of the methylation mark itself. Here, we presented a detailed analysis of seven representative experimental complexes between PHD domains and methylated *H3*K4 peptides. Although the general structures are conserved, the PHDs harbor significant differences in sequences. In order to identify important recurrent interactions, we used a free energy decomposition scheme based on molecular dynamics simulations and subsequent analysis by the MMPBSA method [45]. The binding site of the histone peptide is solvent exposed and exhibits significant flexibility in the positioning of the side chains (see for example Figure S1 and S2). An interesting feature is that, despite this diversity, the energetic contributions of the amino acids that anchor the methylated lysine are remarkably similar (Fig. 3). In agreement with experimental studies, as well as with the observed sequence conservation, the conserved βArom Trp emerges as the crucial residue from the PHD for the binding of methylated K4. However, although necessary, the presence of the conserved tryptophan is not a sufficient condition for the recognition of the methylated *H3*K4 by PHDs.

The sequence context of this particular histone PTM imposes other specific constraints on the binding groove. An important feature of the *H3*K4 methylation mark is that it arises in a sequence context rich in positive charges. Indeed, the N-terminal ARTK motif of histone H3 bears three positive charges. As apparent from our energetic analysis, the first four amino acids of histone H3 form a linear interaction motif [80] and their cognate PHDs share common features that make them perfect effectors for the specific recognition of the N-terminal of histone H3.

Recognition sites for charged residue side chains (K4, R2) mediate the strongest interactions and readily dictate the selectivity toward both the site and a particular PTM on both *H3*K4 and *H3*R2. It is interesting to note that certain PHD domains, while selective for methylated *H3*K4, can tolerate methylation of the nearby *H3*R2, while for other domains, modification of *H3*R2 would be very detrimental to binding. While the electrostatic interactions of the N-terminal *H3*A1 are not as sizable as those of R2 and K4, electrostatic stabilization is provided by the PHD L2 loop and the α-ammonium group is frequently involved in direct interactions with acidic residues side chain, in particular at βP^{+4} position. Besides electrostatic anchoring, the methyl group of *H3*A1 and *H3*T3 side chains forms stabilizing van der Waals interactions in the different complexes, as did the side chains of *H3*R2 and *H3*K4.

The MM/PBSA analysis was used as a starting point to devise a simple predictive model that identified several new human PHD domains as likely methylated *H3*K4 binders. This model does not attempt to perform prediction of absolute values of binding affinities. Indeed, the energetic analysis performed here focuses on electrostatic and van der Waals contributions to free energies and several important simplifications (i.e. neglect of entropy, continuum model of solvation, and limitations inherent to a nonpolarizable classical force field) were introduced. However, the MM/PBSA analysis provides a semi-quantitative estimate of interaction strength and, coupled with sufficient experimental information, is a powerful means to identify the essential molecular determinants for complex formation. An important aspect of the model developed is that it considers each interaction independently, rather than summing all contributions to obtain a global free energy, and, as already pointed out [45,81], this is a major element in the predictive ability of the model.

Free energy decomposition analysis based on the Amber MM-GBSA model, followed by machine-learning training on the resulting data, has been used to devise predictive models to identify different peptides (including modified histones) that bind to a given protein domain [81]. However, large amounts of training data are not always available and our study shows that human intervention in the analysis on a restricted amount of data, followed by testing on a small subset of complexes, also allows for the development of reliable predictive models. These models are easy to implement and interpret. Our predictions have been subsequently validated by independent experimental studies in a number of cases. For the PHD domains where no experimental validation exists, their annotation as nuclear proteins involved in chromatin modifications lends support to our predictions. Considering the approximations inherent to our model, and the rather subtle changes that can modulate affinity between low micromolar (the physiological range for the complexes considered here) and 100 micromolar to millimolar (the poor/non binders range), we expect that the predictions may contain some false positives/negatives. However, this is a common occurrence with experimental proteomics methods, as well. The confrontation of different methods is therefore of paramount importance to arrive at the most reliable binding predictions. In that respect, we identified PHD domains as plausible binders for *H3*K4 that were identified by mass spectrometry [8], such as DIO-1, and also several domains that were not identified by the experimental mass spectrometry study. However, some of those, such as KDM7, were validated by independent biophysical methods. Taken together, these data indicate that molecular

dynamics simulations, coupled with force field based energetic analysis such as that provided by MM/PBSA analysis can be used both on a large enough scale and with sufficient reliability, to warrant a place among established proteomics methods. As exemplified by our study, protocols that are not overly costly in computer time and that rely on the analysis of a few experimental complexes for training can be easily implemented and provide data of competitive quality with respect to experiments.

Acknowledgements

This work was supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Međicale (INSERM), and the Université de Strasbourg (UDS). CG was supported by the CNRS, and Ligue Contre le Cancer. Computing time was provided at the French national computing centers by GENCI (Grand Equipement National de Calcul Intensif) and the Meso-Center for High Performance Computing at the University of Strasbourg, France. We thank Sarah Cianferani for useful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2014.09.015.

References

[1] G. Arents, R.W. Burlingame, B.C. Wang, W.E. Love, E.N. Moudrianakis, The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 10148–10152.
[2] K. Luger, T.J. Rechsteiner, A.J. Flaus, M.M. Waye, T.J. Richmond, Characterization of nucleosome core particles containing histone proteins made in bacteria, J. Mol. Biol. 272 (1997) 301–311.
[3] T. Kouzarides, Chromatin modifications and their function, Cell 128 (2007) 693–705.
[4] L.A. Boyer, K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, R. Jaenisch, Polycomb complexes repress developmental regulators in murine embryonic stem cells, Nature 441 (2006) 349–353.
[5] T.Y. Roh, S. Cuddapah, K. Cui, K. Zhao, The genomic landscape of histone modifications in human T cells, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 15782–15787.
[6] A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.
[7] B.D. Strahl, C.D. Allis, The language of covalent histone modifications, Nature 403 (2000) 41–45.
[8] T. Bartke, M. Vermeulen, B. Xhemalce, S.C. Robson, M. Mann, T. Kouzarides, Nucleosome-interacting proteins regulated by DNA and histone methylation, Cell 143 (2010) 470–484.
[9] J.P. Lambert, T. Pawson, A.C. Gingras, Mapping physical interactions within chromatin by proteomic approaches, Proteomics 12 (2012) 1609–1622.
[10] B.A. Liu, B.W. Engelmann, P.D. Nash, High-throughput analysis of peptide-binding modules, Proteomics 12 (2012) 1527–1546.
[11] M. Nikolov, A. Stutzer, K. Mosch, A. Krasauskas, S. Soeroes, H. Stark, H. Urlaub, W. Fischle, Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns, Mol. Cell. Proteomics 10 (2011) (M110.005371-M005110.005371).
[12] H.G. Stunnenberg, M. Vermeulen, Towards cracking the epigenetic code using a combination of high-throughput epigenomics and quantitative mass spectrometry-based proteomics, Bioessays 33 (2011) 547–551.
[13] M. Vermeulen, H.C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K.K. Lee, J.V. Olsen, A.A. Hyman, H.G. Stunnenberg, M. Mann, Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers, Cell 142 (2010) 967–980.
[14] H.C. Eberl, C.G. Spruijt, C.D. Kelstrup, M. Vermeulen, M. Mann, A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics, Mol. Cell 49 (2013) 368–378.
[15] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J.P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Muller, T. Pawson, A.C. Gingras, C.H. Arrowsmith, S. Knapp, Histone recognition and large-scale structural analysis of the human bromodomain family, Cell 149 (2012) 214–231.
[16] X. Shi, I. Kachirskaia, K.L. Walter, J.H. Kuo, A. Lake, F. Davrazou, S.M. Chan, D.G. Martin, I.M. Fingerman, S.D. Briggs, L. Howe, P.J. Utz, T.G. Kutateladze, A.A. Lugovskoy, M.T. Bedford, O. Gozani, Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36, J. Biol. Chem. 282 (2007) 2450–2455.
[17] X. Li, E.A. Foley, S.A. Kawashima, K.R. Molloy, Y. Li, B.T. Chait, T.M. Kapoor, Examining post-translational modification-mediated protein–protein interactions using a chemical proteomics approach, Protein Sci. 22 (2013) 287–295.
[18] X. Li, E.A. Foley, K.R. Molloy, Y. Li, B.T. Chait, T.M. Kapoor, Quantitative chemical proteomics approach to identify post-translational modification-mediated protein–protein interactions, J. Am. Chem. Soc. 134 (2012) 1982–1985.
[19] M. Jessulat, S. Pitre, Y. Gui, M. Hooshyar, K. Omidi, B. Samanfar, I.e.H. Tan, M. Alamgir, J. Green, F. Dehne, A. Golshani, Recent advances in protein–protein interaction prediction: experimental and computational methods, Expert Opin. Drug Discov. 6 (2011) 921–935.
[20] R.S. Stein, W. Wang, The recognition specificity of the CHD1 chromodomain with modified histone H3 peptides, J. Mol. Biol. 406 (2011) 527–541.
[21] N. Li, R.S. Stein, W. He, E. Komives, W. Wang, Identification of methyllysine peptides binding to CBX6 chromodomain in the human proteome, Mol. Cell. Proteomics 12 (2013) 2750–2760.
[22] U. Schindler, H. Beckmann, A.R. Cashmore, HAT3.1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine-rich region, Plant J. 4 (1993) 137–150.
[23] H. Li, S. Ilin, W. Wang, E.M. Duncan, J. Wysocka, C.D. Allis, D.J. Patel, Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF, Nature 442 (2006) 91–95.
[24] J. Wysocka, T. Swigut, H. Xiao, T.A. Milne, S.Y. Kwon, J. Landry, M. Kauer, A.J. Tackett, B.T. Chait, P. Badenhorst, C. Wu, C.D. Allis, A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling, Nature 442 (2006) 86–90.
[25] P.V. Pena, F. Davrazou, X. Shi, K.L. Walter, V.V. Verkhusha, O. Gozani, R. Zhao, T.G. Kutateladze, Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2, Nature 442 (2006) 100–103.
[26] X. Shi, T. Hong, K.L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Pena, F. Lan, M.R. Kaadige, N. Lacoste, C. Cayrou, F. Davrazou, A. Saha, B.R. Cairns, D.E. Ayer, T.G. Kutateladze, Y. Shi, J. Cote, K.F. Chua, O. Gozani, ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression, Nature 442 (2006) 96–99.
[27] F. Lan, R.E. Collins, R. De Cegli, R. Alpatov, J.R. Horton, X. Shi, O. Gozani, X. Cheng, Y. Shi, Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression, Nature 448 (2007) 718–722.
[28] W.W. Tsai, Z. Wang, T.T. Yiu, K.C. Akdemir, W. Xia, S. Winter, C.Y. Tsai, X. Shi, D. Schwarzer, W. Plunkett, B. Aronow, O. Gozani, W. Fischle, M.C. Hung, D.J. Patel, M.C. Barton, TRIM24 links a non-canonical histone signature to breast cancer, Nature 468 (2010) 927–932.
[29] C.A. Musselman, R.E. Mansfield, A.L. Garske, F. Davrazou, A.H. Kwan, S.S. Oliver, H. Oleary, J.M. Denu, J.P. Mackay, T.G. Kutateladze, Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications, Biochem. J. 423 (2009) 179–187.
[30] F. Chignola, M. Gaetani, A. Rebane, T. Org, L. Mollica, C. Zucchelli, A. Spitaleri, V. Mannella, P. Peterson, G. Musco, The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation, Nucleic Acids Res. 37 (2009) 2951–2961.
[31] S. Qin, L. Jin, J. Zhang, L. Liu, P. Ji, M. Wu, J. Wu, Y. Shi, Recognition of unmodified histone H3 by the first PHD finger of bromodomain-PHD finger protein 2 provides insights into the regulation of histone acetyltransferases monocytic leukemic zinc-finger protein (MOZ) and MOZ-related factor (MORF), J. Biol. Chem. 286 (2011) 36944–36955.
[32] S. Iwase, F. Lan, P. Bayliss, L. de la Torre-Ubieta, M. Huarte, H.H. Qi, J.R. Whetstine, A. Bonni, T.M. Roberts, Y. Shi, The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases, Cell 128 (2007) 1077–1088.
[33] K.S. Champagne, N. Saksouk, P.V. Pena, K. Johnson, M. Ullah, X.J. Yang, J. Cote, T.G. Kutateladze, The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide, Proteins 72 (2008) 1371–1376.
[34] P.V. Pena, R.A. Hom, T. Hung, H. Lin, A.J. Kuo, R.P. Wong, O.M. Subach, K.S. Champagne, R. Zhao, V.V. Verkhusha, G. Li, O. Gozani, T.G. Kutateladze, Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor, J. Mol. Biol. 380 (2008) 303–312.
[35] A. Palacios, I.G. Munoz, D. Pantoja-Uceda, M.J. Marcaida, D. Torres, J.M. Martin-Garcia, I. Luque, G. Montoya, F.J. Blanco, Molecular basis of histone H3K4me3 recognition by ING4, J. Biol. Chem. 283 (2008) 15956–15964.
[36] L.A. Baker, C.D. Allis, G.G. Wang, PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks, Mutat. Res. 647 (2008) 3–12.
[37] E.I. Campos, M.Y. Chin, W.H. Kuo, G. Li, Biological functions of the ING family tumor suppressors, Cell. Mol. Life Sci. 61 (2004) 2597–2613.
[38] M. Unoki, K. Kumamoto, C.C. Harris, ING proteins as potential anticancer drug targets, Curr. Drug Targets 10 (2009) 442–454.
[39] S.A. Jacobs, S. Khorasanizadeh, Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail, Science 295 (2002) 2080–2083.
[40] R.M. Hughes, K.R. Wiggins, S. Khorasanizadeh, M.L. Waters, Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 11184–11188.
[41] Z. Lu, J. Lai, Y. Zhang, Importance of charge independent effects in readout of the trimethyllysine mark by HP1 chromodomain, J. Am. Chem. Soc. 131 (2009) 14928–14931.
[42] S.D. Taverna, H. Li, A.J. Ruthenburg, C.D. Allis, D.J. Patel, How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers, Nat. Struct. Mol. Biol. 14 (2007) 1025–1040.
[43] D. Spiliotopoulos, A. Spitaleri, G. Musco, Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study, PLoS One 7 (2012) (e46902-e46902).
[44] M. Ozboyaci, A. Gursoy, B. Erman, O. Keskin, Molecular recognition of H3/H4 histone tails by the tudor domains of JMJD2A: a comparative molecular dynamics simulations study, PLoS One 6 (2011) (e14765-e14765).

[45] V. Lafont, M. Schaefer, R.H. Stote, D. Altschuh, A. Dejaegere, Protein–protein recognition and interaction hot spots in an antigen–antibody complex: free energy decomposition identifies "efficient amino acids", Proteins 67 (2007) 418–434.
[46] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897.
[47] G.G. Wang, J. Song, Z. Wang, H.L. Dormann, F. Casadio, H. Li, J.L. Luo, D.J. Patel, C.D. Allis, Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger, Nature 459 (2009) 847–851.
[48] Z. Wang, J. Song, T.A. Milne, G.G. Wang, H. Li, C.D. Allis, D.J. Patel, Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression, Cell 141 (2010) 1183–1194.
[49] H. Wen, J. Li, T. Song, M. Lu, P.Y. Kan, M.G. Lee, B. Sha, X. Shi, Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation, J. Biol. Chem. 285 (2010) 9322–9326.
[50] J.R. Horton, A.K. Upadhyay, H.H. Qi, X. Zhang, Y. Shi, X. Cheng, Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases, Nat. Struct. Mol. Biol. 17 (2010) 38–43.
[51] M. Fiedler, M.J. Sanchez-Barrena, M. Nekrasov, J. Mieszczanek, V. Rybin, J. Muller, P. Evans, M. Bienz, Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex, Mol. Cell 30 (2008) 507–518.
[52] S. Ramon-Maiques, A.J. Kuo, D. Carney, A.G. Matthews, M.A. Oettinger, O. Gozani, W. Yang, The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 18993–18998.
[53] H. van Ingen, F.M. van Schaik, H. Wienk, J. Ballering, H. Rehmann, A.C. Dechesne, J.A. Kruijzer, R.M. Liskamp, H.T. Timmers, R. Boelens, Structural insight into the recognition of the *H3*K4me3 mark by the TFIID subunit TAF3, Structure 16 (2008) 1245–1256.
[54] S.D. Taverna, S. Ilin, R.S. Rogers, J.C. Tanny, H. Lavender, H. Li, L. Baker, J. Boyle, L.P. Blair, B.T. Chait, D.J. Patel, J.D. Aitchison, A.J. Tackett, C.D. Allis, Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs, Mol. Cell 24 (2006) 785–796.
[55] H.A. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[56] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pKa values, Proteins 61 (2005) 704–721.
[57] A.T. Brunger, M. Karplus, Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison, Proteins 4 (1988) 148–156.
[58] A.D. MacKerell, D. Bashford, M. Bellot, R.L.D. Jr., J.D. Evanseck, M.J.F., S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E.R. Iii, B. Roux, M. schlenkrich, J.C. smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical force field for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[59] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[60] C. Grauffel, R.H. Stote, A. Dejaegere, Force field parameters for the simulation of modified histone tails, J. Comput. Chem. 31 (2010) 2434–2451.
[61] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[62] M.E. Davis, J.D. Madura, B.A. Luty, J.A. McCammon, Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian, Dynamics Program, Comput. Phys. Commun. 62 (1991) 187–197.
[63] V. Zoete, O. Michielin, M. Karplus, Protein-ligand binding free energy estimation using molecular mechanics and continuum electrostatics. Application to HIV-1 protease inhibitors, J. Comput. Aided Mol. Des. 17 (2003) 861–880.
[64] S. Ren, G. Yang, Y. He, Y. Wang, Y. Li, Z. Chen, The conservation pattern of short linear motifs is highly correlated with the function of interacting protein domains, BMC Genomics 9 (2008) 452.
[65] A. Bateman, L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats, S.R. Eddy, The Pfam protein families database, Nucleic Acids Res. 32 (2004) D138–D141.
[66] N. authors listed,. Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[67] Y. Tsukada, J. Fang, H. Erdjument-Bromage, M.E. Warren, C.H. Borchers, P. Tempst, Y. Zhang, Histone demethylation by a family of JmjC domain-containing proteins, Nature 439 (2006) 811–816.
[68] A.M. Rojas, L. Sanchez-Pulido, A. Futterer, K.H. van Wely, C. Martinez-A, A. Valencia, Death inducer obliterator protein 1 in the context of DNA regulation. Sequence analyses of distant homologues point to a novel functional role, FEBS J. 272 (2005) 3505–3511.
[69] J. Gatchalian, A. Futterer, S.B. Rothbart, Q. Tong, H. Rincon-Arano, A. Sanchez de Diego, M. Groudine, B.D. Strahl, C. Martinez-A, K.H. van Wely, T.G. Kutateladze, Dido3 PHD modulates cell differentiation and division, Cell Rep. 4 (2013) 148–158.
[70] S. Braig, A.K. Bosserhoff, Death inducer–obliterator 1 (Dido1) is a BMP target gene and promotes BMP-induced melanoma progression, Oncogene 32 (2013) 837–848.
[71] C.M. Santiveri, M.F. Garcia-Mayoral, J.M. Perez-Canadillas, M.A. Jimenez, NMR structure note: PHD domain from death inducer obliterator protein and its interaction with H3K4me3, J. Biomol. NMR 56 (2013) 183–190.
[72] G. Cui, S. Park, A.I. Badeaux, D. Kim, J. Lee, J.R. Thompson, F. Yan, S. Kaneko, Z. Yuan, M.V. Botuyan, M.T. Bedford, J.Q. Cheng, G. Mer, PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53, Nat. Struct. Mol. Biol. 19 (2012) 916–924.
[73] S. An, K.J. Yeo, Y.H. Jeon, J.J. Song, Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism, J. Biol. Chem. 286 (2011) 8369–8374.
[74] D.H. Ku, C.D. Chang, J. Koniecki, L.A. Cannizzaro, L. Boghosian-Sell, H. Alder, R. Baserga, A new growth-regulated complementary DNA with the sequence of a putative trans-activating factor, Cell Growth Differ. 2 (1991) 179–186.
[75] J. Cheng, Y. Yang, J. Fang, J. Xiao, T. Zhu, F. Chen, P. Wang, Z. Li, H. Yang, Y. Xu, Structural insight into coordinated recognition of trimethylated histone H3 Lysine 9 (H3K9me3) by the Plant Homeodomain (PHD) and Tandem Tudor Domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein, J. Biol. Chem. 288 (2013) 1329–1339.
[76] Y.J. Kim, H. Young Kim, J.H. Lee, S. Jong Yu, J.H. Yoon, H.S. Lee, C. Yong Kim, J. Youn Cheong, S. Won Cho, N. Hwa Park, B. Lae Park, S. Namgoong, L. Hyo Kim, H. Sub Cheong, H. Doo Shin, A genome-wide association study identified new variants associated with the risk of chronic hepatitis B, Hum. Mol. Genet. 22 (2013) 4233–4238.
[77] U. Fischer, A.K. Struss, D. Hemmer, A. Michel, W. Henn, W.I. Steudel, E. Meese, PHF3 expression is frequently reduced in glioma, Cytogenet. Cell Genet. 94 (2001) 131–136.
[78] L. Zuo, X. Zhang, H.W. Deng, X. Luo, Association of rare PTP4A1-PHF3-EYS variants with alcohol dependence, J. Hum. Genet. 58 (2013) 178–179.
[79] L.E. Machado, Y. Pustovalova, A.C. Kile, A. Pozhidaeva, K.A. Cimprich, F.C. Almeida, I. Bezsonova, D.M. Korzhnev, PHD domain from human SHPRH, J. Biomol. NMR 56 (2013) 393–399.
[80] N.E. Davey, K. Van Roey, R.J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, T.J. Gibson, Attributes of short linear motifs, Mol. BioSyst. 8 (2012) 268–281.
[81] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, J. Chem. Inf. Model. 51 (2011) 69–82.