3448–3463 *Nucleic Acids Research, 2016, Vol. 44, No. 7* *Published online 20 February 2016*
doi: 10.1093/nar/gkw103

Structural basis for PHD_VC5HCH_{NSD1}–$C2HR_{Nizp1}$ interaction: implications for Sotos syndrome

Andrea Berardi[1,2,†], Giacomo Quilici[1,†], Dimitrios Spiliotopoulos[1,3,†], Maria Angeles Corral-Rodriguez[1,3], Fernando Martin-Garcia[1], Massimo Degano[4], Giovanni Tonon[5], Michela Ghitti[1] and Giovanna Musco[1,*]

[1]Biomolecular NMR Unit, Division of Genetics and Cell Biology, IRCCS S. Raffaele Scientific Institute, Milan 20132, Italy, [2]Università degli Studi di Milano, Italy, [3]Università Vita e Salute San Raffaele, Milano 21032, Italy, [4]Biocrystallography Unit, Division of Immunology, Transplantation, and Infectious Diseases, IRCCS S. Raffaele Scientific Institute, Milan 20132, Italy and [5]Functional genomics of cancer, Division of Experimental Oncology, IRCCS S. Raffaele Scientific Institute, Milan 20132, Italy

Received December 28, 2015; Revised February 03, 2016; Accepted February 09, 2016

ABSTRACT

Sotos syndrome is an overgrowth syndrome caused by mutations within the functional domains of *NSD1* gene coding for NSD1, a multidomain protein regulating chromatin structure and gene expression. In particular, PHD_VC5HCH_{NSD1} tandem domain, composed by a classical (PHD_V) and an atypical (C5HCH) plant homeo-domain (PHD) finger, is target of several pathological missense-mutations. PHD_VC5HCH_{NSD1} is also crucial for NSD1-dependent transcriptional regulation and interacts with the C2HR domain of transcriptional repressor Nizp1 ($C2HR_{Nizp1}$) *in vitro*. To get molecular insights into the mechanisms dictating the patho-physiological relevance of the PHD finger tandem domain, we solved its solution structure and provided a structural rationale for the effects of seven Sotos syndrome point-mutations. To investigate PHD_VC5HCH_{NSD1} role as structural platform for multiple interactions, we characterized its binding to histone H3 peptides and to $C2HR_{Nizp1}$ by ITC and NMR. We observed only very weak electrostatic interactions with histone H3 N-terminal tails, conversely we proved specific binding to $C2HR_{Nizp1}$. We solved $C2HR_{Nizp1}$ solution structure and generated a 3D model of the complex, corroborated by site-directed mutagenesis. We suggest a mechanistic scenario where NSD1 interactions with cofactors such as Nizp1 are impaired by PHD_VC5HCH_{NSD1} pathological mutations, thus impacting on the repression of growth-promoting genes, leading to overgrowth conditions.

INTRODUCTION

The Nuclear receptor-binding SET (Su(var) 3–9, Enhancer of zeste, Trithorax) domain protein 1 (NSD1), is a large multi-domain nuclear protein (2696 amino acids) belonging to a family of three structurally similar mammalian methyltransferases including NSD1, NSD2 and NSD3, that have been all linked to multiple diseases and cancer types (1–4). As NSD prototype, NSD1 is characterized by the presence of different chromatin related domains including two proline–tryptophan–tryptophan–proline domains (PWWP), five plant homeodomains (PHD), a PHD finger-like Cys–His rich domain (C5HCH) (sometimes called PHD_{VI}), acquired late in the evolution of the NSD family and a catalytic SET domain (5–7). *In vitro* NSD1-SET domain catalyzes the mono- and di-methylation of H3K36 and tri-methylation of H4K20, even though discrepancies exist as to the level and the target(s) of methylation (8–13). Importantly, mutations in the *NSD1* gene lead to different aberrant developmental processes: fusion to *Nup98* nucleoporin gene is associated with *de novo* childhood acute myeloid leukemia (AML) (14–16), and *NSD1* mutations and translocations lead to Weaver and Sotos syndromes, inherited congenital malformation overgrowth syndromes leading to delayed motor and cognitive development (17–21). Indeed, mouse *NSD1* knock out data show that the protein is essential for correct embryonic development (6). Although the different patho-physiological mechanisms dictated by NSD1 remain elusive, several lines of evidence suggest its direct involvement in context-dependent transcriptional repression or activation. In colon cancer cell

*To whom correspondence should be addressed. Tel: +39 0226434824; Fax: +39 0226434153; Email: musco.giovanna@hsr.it
†The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.
Present address: Dimitrios Spiliotopoulos, Dimitrios Spiliotopoulos, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

lines NSD1 binds near various promoter elements tuning the levels of the various H3K36 methylation forms within the occupied promoter proximal region, regulating multiple genes involved in developmental processes, such as cell growth/cancer and bone morphogenesis (9). Conversely, in neuroblastoma cells NSD1 displays tumour suppressor like properties promoting *MEIS1*-mediated gene transcriptional repression (8). NSD1 bi-functional transcriptional regulation might be also related not only to the presence of two nuclear-interaction domains, NID^{-L} and NID^{+L} (5) but also to the presence of the other chromatin binding domains, that are expected to confer to NSD1 additional chromatin related functions, besides its SET-dependent methyltransferase activity. As a matter of fact, the fifth PHD finger and the adjacent C5HCH domain (PHD_VC5HCH_{NSD1}) have been shown to have context-dependent activation or repression activities with different patho-physiological outcomes. On the one hand, PHD_VC5HCH_{NSD1} is a hot spot for the developmental overgrowth Sotos syndrome, as assessed by the presence of 17 pathological point-mutations within this domain (Supplementary Table S1). On the other hand, in AML context the same tandem domain contributes to inappropriate *Hoxa7* and *HoxA9* genes activation (22). This is in keeping with the chromatin associated functions generally attributed to these evolutionarily conserved Zn^{2+} binding 'reader/effector' modules (~60 aminoacids long). PHD fingers usually interpret histones post-translational modifications (H3K4 *vs.* H3K4me3/2; H3R2me0 *vs.* H3R2me2; H3K36me3; H3K14ac) in a modification and context-specific fashion, thus promoting chromatin changes and/or protein recruitment (23,24). Consistent with its putative role as chromatin reader, murine PHD_VC5HCH_{NSD1} (99% identity with human PHD_VC5HCH_{NSD1} (Supplementary Figure S1) binds biotinylated $H3K4me3_{1-21}$ and $H3K9me3_{1-21}$ peptides *in vitro* (25), even though this interaction has been recently challenged (26). In fact, GST-pulldown assays using unfractionated calf-thymus histones or biotinylated histone peptides did not prove evidence of histone binding to PHD_VC5HCH_{NSD1}, thus raising a conflict in literature about its actual role as epigenetic reader (26). Notably, PHD fingers are emerging as a robust-conserved structural scaffold working as versatile non-histone binding domains, thereby extending their role to diverse cellular processes, far beyond the well documented histone tail interpretation (27,28). In line with the multifaceted role of PHD fingers, PHD_VC5HCH_{NSD1} seems to function as a hub for the interaction with other proteins/domains critical for transcriptional activity, such as the C2HR domain of the transcriptional repressor Nizp1 (NSD1 interacting Zinc-finger protein), one of the few documented NSD1 interactors (7,25,26,29). Nizp1, is a poorly characterized multidomain protein, expressed in several tissues containing an N-terminal SCAN box, a repressor KRAB domain, an atypical C2HR Zinc-finger motif ($C2HR_{Nizp1}$) followed by four classical C2H2-type Zinc-fingers (7,29). Intriguingly, according to biochemical *in-vitro* experiments the interaction with $C2HR_{Nizp1}$ seems to be a unique peculiarity of PHD_VC5HCH_{NSD1} (7,26), thus implying a functional divergence within the NSD protein family. In order to move a step further in the comprehension of the molecular mechanisms dictating PHD_VC5HCH_{NSD1} patho-physiological relevance, we solved its NMR solution structure and provided also a structural rationale for the effects of seven Sotos syndrome point-mutations. To investigate the potential role of PHD_VC5HCH_{NSD1} as structural platform for multiple interactions we characterized its binding to histone H3 peptides and to $C2HR_{Nizp1}$ by ITC and NMR. We observed only very weak electrostatic interactions with histone H3 N-terminal tail, conversely we proved the existence of a specific interaction ($K_d \sim 4\ \mu M$) between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$. We also solved the solution structure of $C2HR_{Nizp1}$ and generated a three-dimensional model of its complex with PHD_VC5HCH_{NSD1} corroborated by site-directed mutagenesis studies. We suggest a mechanistic scenario where NSD1 interactions with cofactors, such as Nizp1 are impaired by PHD_VC5HCH_{NSD1} pathological mutations, thus impacting on the repression of growth-promoting genes, and ultimately leading to overgrowth conditions.

MATERIALS AND METHODS

Sample preparation for NMR and binding assays

Murine NSD1 PHD_vC5HCH (Glu2117–Asp2207), (corresponding to residues Glu2116–Asp2206 in the human sequence NM_022455.4) was amplified from a pSG5 vector containing the cDNA for murine NSD1 (NCBI Reference Sequence: NM_008739.3). Murine $C2HR_{Nizp1}$ (residues Glu397–Lys434) was amplified from a pCVM vector containing the cDNA coding for mouse Nizp1 (protein NCBI Reference Sequence: NM_032752.1). Both plasmids were kindly provided by Prof. Chambon (Strasbourg, France). pETM11 and pETM41 (EMBL) plasmids were used to express His-tagged-PHD_vC5HCH_{NSD1} and His-MBP-tagged-$C2HR_{Nizp1}$ proteins, respectively. The tags were removed by cleavage with TEV-protease. Site-directed mutagenesis was performed by standard overlap extension methods. The DNA constructs were sequenced by Eurofins (Milan, Italy). Both protein domains were expressed in *Escherichia coli* BL21 (DE3) cells at 28°C overnight after induction with 1mM isopropyl thio-β-D-galactoside (IPTG), in LB medium supplemented with 0.2 mM $ZnCl_2$. Uniformly ^{15}N- and ^{13}C-^{15}N-labeled PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$ were expressed by growing *E. coli* BL21 (DE3) cells in minimal bacterial medium containing $^{15}NH_4Cl$, with or without ^{13}C-D-glucose as sole nitrogen and carbon sources. Proteins were purified as described in (30). For binding assays with histone peptide arrays PHD_vC5HCH_{NSD1} was cloned into pETM30 expression vector (EMBL) containing an N-terminal His-GST tag. As control only His-GST was used. The molecular masses of the recombinant proteins were checked by mass spectrometry (MALDI). Synthetic histone H3 peptides ($H3_{1-10}$, $H3_{1-21}$, $H3_{1-37}$, $H3K4me3_{1-21}$, $H3K9me3_{1-21}$) were C-amidated. $C2HR_{Nizp1}$ peptides used for ITC and NMR titrations were N-acetylated and C-amidated. They were purchased from Caslo Lyngby, Denmark. Peptide purity (>98%) was confirmed by HPLC and mass spectrometry. The NMR buffer of both PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$ contained 20 mM sodium phosphate pH 6.3, 0.15 M NaCl, 2 mM dithiothreitol (DTT), 50 μM $ZnCl_2$

(28,31) with 0.15 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). D_2O was 10% (v/v) or 100% depending on the experiments.

NMR spectroscopy and resonance assignment

NMR experiments were performed at 295 K on a Bruker Avance 600 MHz equipped with inverse triple-resonance cryoprobe and pulsed field gradients (Bruker, Karlsruhe, Germany). Typical sample concentration was 0.3–0.4 mM. Data were processed using NMRPipe (32) or Topspin 3.2 (Bruker) and analyzed with CCPNmr Analysis 2.1 (33). The ^{1}H, ^{13}C, ^{15}N chemical shifts of the backbone atoms of both PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$ were obtained from three-dimensional HNCA, CBCA(CO)NH, CBCANH, HNCO experiments. Side chain resonances were assigned through H(CCO)NH, CC(CO)NH, HCCH–TOCSY experiments (34) and classical ^{1}H–^{1}H 2D-TOCSY (Total Correlation Spectroscopy). The resonances of the aromatic side chains atoms were assigned using ^{1}H–^{13}C HSQC, (HB)CB(CGCD)HD and the 2D TOCSY and NOESY (Nuclear Overhauser effect spectroscopy) experiments in H_2O and in D_2O. The tautomeric state of the imidazole rings of the histidines were determined performing 2D ^{1}H–^{15}N HMQC (Heteronuclear Multiple Quantum Coherence) long-range experiments with transfer delay optimized to transfer the coherence from the Hϵ1 and Hδ2 atoms to the Nϵ2 and Nδ1 atoms (τ = 11 ms) and analyzing the peaks pattern according to (35). Proton–proton distance constraints were obtained from ^{15}N and ^{13}C separated three-dimensional NOESY spectra, 2D ^{1}H–^{1}H NOESY in D_2O and H_2O employing 100 ms mixing times. φ/ψ restraints were obtained from backbone chemical shifts using TALOS-N (36). Hydrogen bond restraints were defined from slow-exchanging amide protons identified after exchange in D_2O. ^{1}H–^{15}N residual dipolar couplings were measured in isotropic and anisotropic phases generated by an axially compressed acrylamide gel (8%) as described in (37).

Relaxation experiments

Heteronuclear $\{^{1}H\}$ ^{15}N nuclear Overhauser enhancement, longitudinal and transversal ^{15}N relaxation rates (R1, R2) were measured using standard 2D methods (38). For R1 and R2 experiments a duty-cycle heating compensation was used (39), the two decay curves were sampled at 13 (ranging from 50 to 2000 ms) and 11 (from 12 to 244 ms) different time points, respectively. Both R1 and R2 experiments were collected in random order and using a recovery delay of 2.5 s. R1 and R2 values were fitted to a 2-parameter exponential decay from the intensities using the fitting routine implemented in the analysis program NMRView (40), duplicate measurements are used to determine the uncertainty in the relaxation rate by Monte Carlo methods in NMRView. The tumbling correlation time (τ_c) was estimated from R2/R1. Only residues corresponding to non-overlapping peaks, with heteronuclear NOE > 0.65 were used, residues with anomalous short transverse relaxation time T2 were excluded (41). HSQC spectra measured in an interleaved fashion with and without 4 seconds of proton saturation during recovery delay were recorded for the $\{^{1}H\}$–^{15}N heteronuclear NOE experiments. The corresponding values were obtained from the ratio between saturated and unsaturated peaks intensities. The uncertainty was calculated as the standard deviation of the noise in the spectrum divided by the intensity of the reference peak.

Structure calculations and validation

PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$ structures were calculated with ARIA 2.3.2 (42) in combination with CNS using experimentally derived restraints (Table 1). All NOEs were assigned manually and calibrated by ARIA, the automated assignment was not used. A total of eight iterations was performed, computing 20 structures in the first seven iterations and 200 in the last iteration. The ARIA default water refinement was performed on the 40 best structures of the final iteration. Initial structures were calculated without Zn^{2+} restraints to verify the position and the geometry of the metal ion ligands, so that the residues involved in Zn^{2+} binding could be identified in an unbiased manner. Several NOEs were observed between metal coordinating residues. After unequivocal identification of the metal coordinating residues, in final refinement calculations the geometry of the Zn^{2+} coordination was fixed through covalent bonds and angles in the CNS parameters ($d_{Zn\text{-}S\gamma}$ = 2.3 Å; $d_{Zn\text{-}N\epsilon2,N\delta1}$ = 2.0 Å tetrahedral angle geometry around Zn^{2+} ion, i.e. Sγ–Zn–Sγ = 109.5°, Sγ–Zn–(Nδ1 or Nϵ2) = 109.5°); the proper angles and distances for Zn^{2+} ions coordinating residues were maintained also during water refinement. PROCHECK-NMR (43) and iCING (44) were used to assess the structural quality. The family of the 15 lowest energy structures (no distance or torsional angle restraints violations >0.5 Å or >5°, respectively) of PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$ have been deposited in the PDB (PDB accession codes 2naa and 2nab, respectively). Chemical shift and restraints lists used for structure calculations have been deposited in BioMagResBank (PHD_vC5HCH_{NSD1} accession code: 25933; $C2HR_{Nizp1}$ accession code 25934). Electrostatic potential maps were calculated using APBS plugin of PyMOL (45). Figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

NMR binding assays

Protein concentrations were determined by UV spectroscopy using the predicted extinction coefficients of 12490 and 6990 $M^{-1}cm^{-1}$ for PHD_vC5HCH_{NSD1} and $C2HR_{Nizp1}$, respectively. Peptide concentrations were estimated from their dry-weight. In order to minimize dilution and NMR signal loss, titrations were carried out by adding to the ^{15}N labeled protein samples (typically 0.2–0.3 mM) small aliquots of concentrated (15 mM) peptides or proteins (1–2 mM) stock solutions. All the solutions were prepared in NMR buffer. For each titration point (stepwise addition of up to 12 and 2 equivalents for histone H3 peptides and for PHD_vC5HCH_{NSD1}/$C2HR_{Nizp1}$ titrations, respectively) a 2D water-flip-back ^{15}N-edited HSQC spectrum was acquired with 512 (100) complex points, 55 ms (60 ms) acquisition times, apodized by 60° shifted squared (sine) win-

dow functions and zero filled to 1024 (512) points for 1H (15N), respectively. Assignment of the PHD_vC5HC_{NSD1} in the presence of histone peptides was made by following individual cross-peaks through the titration series. The assignments of PHD_VC5HCH_{NSD1} and of $C2HR_{Nizp1}$ amides were achieved acquiring HNCA, CBCA(CO)NH experiments in the presence of a 1.5 fold excess of unlabelled $CH2R_{Nizp1}$ and unlabelled PHD_VC5HCH_{NSD1}, respectively. For each residue the weighted average of the ^1H and ^{15}N chemical shift perturbation (CSP) was calculated as $CSP = [(\Delta\delta^2HN + \Delta\delta^2N/25)/2]^{1/2}$ (46).

Isothermal titration calorimetry thermodynamic analysis

ITC titration was performed using a VP-ITC isothermal titration calorimeter (MicroCal LLC, Northampton, MA, USA). Recombinant proteins and synthetic peptides were dialyzed against the same buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 7.2, 150 mM NaCl, 2 mM β-mercaptoethanol, 10 μM $ZnCl_2$) at 23°C. Step by step injections of the 0.3–10 mM titrants ($H3K9me3_{1-21}$, $C2HR_{Nizp1}$ and mutants) solution into a cell containing a 25–200 μM PHD_VC5HCH_{NSD1} were performed to finally reach a 2-fold and 7-fold molar excess of $CH2R_{Nizp1}$ and $H3K9me3_{1-21}$, respectively, with respect to the protein concentration. The quantity of heat absorbed or released in the process was measured. Control experiments were performed under identical conditions to determine the dilution heat of the titrant peptides into buffer and of the buffer into protein samples. The final data were analyzed using the software ORIGIN 7.0 (OriginLab Corp., Northampton, MA, USA).

Quantum mechanics/molecular mechanics calculations

$C2HR_{Nizp1}$ structure calculations suggested that the Zn^{2+} ion was coordinated only by the thiols of $C407_{Nizp1}$ and $C410_{Nizp1}$ and by the Nϵ2 of $H423_{Nizp1}$, we therefore verified whether water molecules could complete the Zn^{2+} coordination sphere using hybrid quantum mechanics/molecular mechanics (QM/MM) calculations (Density Functional based Tight Binding) (47). Calculations have been performed on the final structure of a MD simulation (100 ns) performed on the lowest $C2HR_{Nizp1}$ NMR energy structure using Amber14 (48), the Amber ff12SB force field and the SPCE water model (49). Adaptive solvent QM/MM calculations (50) were performed including in the active region the side chains of $C407_{Nizp1}$, $C410_{Nizp1}$, $H423_{Nizp1}$, $N409_{Nizp1}$ and eleven water molecules. After the equilibration phase of the QM part, a free QM/MM simulation (100 ps) was performed using S*ander* module (48).

Docking model

The complex model was generated using the information-driven docking software HADDOCK 2.0 (51). 10 NMR input structures for both PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ (residues S401-E429) were used. Ambiguous interactions restraints (AIRs) were generated based on NMR CSP data. We defined as *active* those residues showing a significant amide chemical shift displacement (defined as CSP > <CSP> + 1σ, whereby σ is the standard deviation) upon ligand binding and as *passive* those residues within a radius of 4 Å from the active ones. Among them, only amino acids with high solvent accessibility (relative all-atoms accessibilities > 40%) were considered (Supplementary Table S2). Naccess V2.1.1 (Hubbard, S.J. & Thornton, J.M. (1993), 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London) was used for the solvent accessible area calculations. Unambiguous restraints were applied to maintain the Zn^{2+} ion coordination sphere. Calculations generated 1000, 1000, 500 structures for the rigid body docking (it0), the semi-flexible refinement (it1) and the explicit solvent refinement (water), respectively. In it0, each model was the result of five internal docking trials (with randomization of the starting orientations), whereby for each model the pose rotated by 180° around the normal to the interface was also sampled. In the it1 step, the semi-flexible regions were automatically defined. OPLS force field (52) and TIP3P water model were used. Tetra-coordinated Zn^{2+} ions in PHD_VC5HCH_{NSD1} were treated according to bonded ZAFF parameters (53), whereas the Zn^{2+} ion in $C2HR_{Nizp1}$ was treated according to the aforementioned QM/MM calculations. The final 500 structures obtained after water refinement were scored according to their Haddock Score. The latter (defined as $HADDOCK_{score}$ = 1.0 EvdW+0.2 Eelec+1.0 Edesolv+0.1 EAIR) is a weighted combination of van der Waals and electrostatic energy terms, represented by Lennard–Jones and Coulomb potentials, empirical desolvation term developed by Recio *et al.* (54) and ambiguous interaction restraint energy term, which reflects the accordance of the model to the input restraints. HADDOCK models were clustered (55) based on their interface root mean square deviation (r.m.s.d.), setting the cutoff and the minimum number of models in a cluster to 4 Å and to 10, respectively. To remove any bias of the cluster size on the cluster statistics, the final overall score of each cluster was calculated on the five lowest HADDOCK scores models in that cluster (Supplementary Table S3).

RESULTS AND DISCUSSION

PHD_VC5HCH_{NSD1} tandem domain forms a unique structural entity

We determined the three-dimensional solution structure of PHD_VC5HCH_{NSD1} using multidimensional heteronuclear NMR spectroscopy (Figure 1A, B, Table 1). As the mouse PHD_VC5HCH_{NSD1} sequence shares 99% identity with the human one (Supplementary Figure S1), we adopted the human sequence numbering scheme to map Sotos syndrome mutants (see further). Residues $E2116\text{-}Q2163_{NSD1}$ adopt the classical PHD finger fold consisting of an initial loop, relatively well defined from residues $D2126_{NSD1}$ to $L2130_{NSD1}$, a two-stranded antiparallel β-sheet formed by $L2130\text{-}S2132_{NSD1}$ (β1) and $V2141\text{-}H2143_{NSD1}$ (β2), followed by a short α-helix ($A2144\text{-}L2147_{NSD1}$), and the so-called variable L3 loop ($N2148\text{-}E2158_{NSD1}$) (56). A C-terminal 3–10 α-helix ($P2160\text{-}H2162_{NSD1}$) links the domain to the adjacent C5HCH domain (residues $C2164\text{-}D2206_{NSD1}$), whose sole elements of secondary structures consist of two short antiparallel β-sheets, β3-β4 ($S2173\text{-}C2175_{NSD1}$; $S2180\text{-}F2182_{NSD1}$), β5-β6

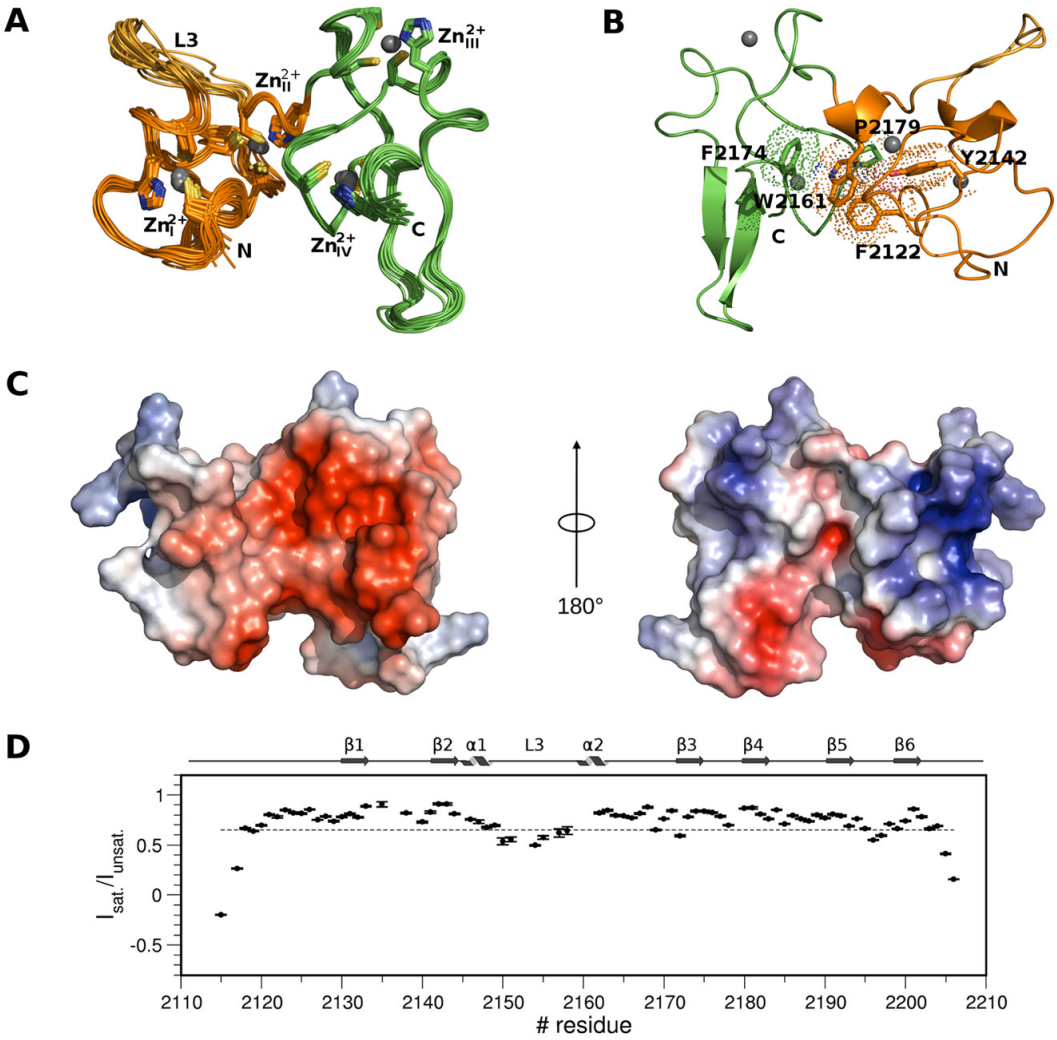

Figure 1. Solution structure of PHD_VC5HCH_{NSD1}. (**A**) Superposition of the best 15 NMR structures PHD_V; the L3 loop and C5HCH are coloured in orange, gold and green, respectively. Zn^{2+} binding residues and Zn^{2+} ions are represented in sticks and spheres, respectively. (**B**) Cartoon representation of PHD_VC5HCH_{NSD1} highlighting elements of secondary structure. Residues forming the domain hydrophobic core at the domains interface are shown in sticks and dotted space-filling representation. Zn^{2+} ions are represented in spheres. Amino acids in this and in the following Figures are numbered according to the human NSD1 sequence. (**C**) Electrostatic surface potential of PHD_VC5HCH_{NSD1}. The structures on the left and on the right are represented in the same orientation as in (A) and (B), respectively. (**D**) Backbone dynamics of PHD_VC5HCH_{NSD1}. Dotted line indicates the $\{^1H\}$–^{15}N heteronuclear NOE value threshold of 0.65. Elements of secondary structure are indicated on the top.

($L2191\text{-}S2194_{NSD1}$; $G2198\text{-}C2202_{NSD1}$) arranged orthogonally one to each other (Figure 1A and B). The tandem domain is stabilized by four Zn^{2+} ions, coordinated respectively by $C2121_{NSD1}$, $C2124_{NSD1}$, $C2146_{NSD1}$ and $H2143_{NSD1}$ ($Zn^{2+}{}_I$), $C2133_{NSD1}$, $C2138_{NSD1}$, $C2159_{NSD1}$ and $H2162_{NSD1}$ ($Zn^{2+}{}_{II}$), $C2164_{NSD1}$, $C2167_{NSD1}$, $C2183_{NSD1}$ and $H2186_{NSD1}$ ($Zn^{2+}{}_{III}$), $C2175_{NSD1}$, $C2178_{NSD1}$, $C2202_{NSD1}$ and $H2205_{NSD1}$ ($Zn^{2+}{}_{IV}$) (Figure 1A). Metal ions binding via the imidazole rings occurs through Nδ of $H2143_{NSD1}$ and $H2186_{NSD1}$ and through Nϵ atoms for $His2162_{NSD1}$ and $H2205_{NSD1}$, as assessed from HMQC long range experiments (data not shown). Several conserved hydrophobic residues within the NSD family ($F2122_{NSD1}$,$Y2142_{NSD1}$, $W2161_{NSD1}$, $F2174_{NSD1}$, $P2179_{NSD1}$) (Supplementary Figure S1A) establish direct inter-domain interactions (Figure 1B) creating a hydrophobic groove surrounded by negatively charged residues (Figure 1C). Of note, when expressed as single domains both the PHD_V and the C5HCH domains are unstable in solution, as assessed respectively, by their low solubility and NMR peaks duplication (data not shown), thus supporting the notion that the two Zn^{2+} binding domains are intimately linked, forming an indivisible structural entity. Global and local dynamics of the tandem domain, including motions ranging from picoseconds to milliseconds time scales, were obtained analyzing ^{15}N relaxation parameters from 79 residues that gave sufficiently resolved backbone peaks in the ^1H-^{15}N HSQC spectrum. The R2/R1 ratio for residues located in both PHD_V and C5HCH is relatively uniform indicating that in solution the molecule behaves as a monomer and tumbles as a single body with a calculated correlation time of 7.96±0.41 ns (Supplementary Figure S2A). The steady-state heteronuclear $\{^1H\}$–^{15}N NOE, a sensitive indicator of internal motions on a

Table 1. Summary of conformational constraints and statistics for the 15 best structures of PHD_VC5HCH_{NSD1}

Restraints information[a,b]	<SA>[c]
Total number of experimental distance restraints	1344
NOEs (intraresidual/sequential/medium/long)	631/273/137/303
Zn^{2+} coordination restraints	16
Hydrogen bonds	7
Dihedral angle restraints (Φ/Ψ)	57/57
Residual dipolar couplings[d]	22
Deviation from idealized covalent geometry	
Bonds (Å)	0.0031 ± 0.0001
Angles (°)	0.498 ± 0.012
Coordinate r.m.s. deviation (Å)[e]	
Ordered backbone atoms (N, Cα, C')	0.67 ± 0.09
Ordered heavy atoms	1.06 ± 0.09
Ramachandran quality parameters (%)[e]	
Residues in most favored regions	89.40%
Residues in allowed regions	10.60%

[a] No distance restraint in any of the structures included in the ensemble was violated by more then 0.5 Å.
[b] No dihedral angle restraints in any of the structures included in the ensemble was violated by more than 5°.
[c] Simulated annealing, statistics refers to the ensemble of 15 structures with the lowest energy.
[d] RDC restraints have been used only for residues adopting regular secondary structure.
[e] Root mean square deviation between the ensemble of structures <SA> and the lowest energy structure and Ramachandran quality parameters calculated on residues Asp2119–Asp2205.

sub-nanosecond timescale, revealed that most residues displayed values around 0.8. The N- and C-termini, the L3-loop within PHD_V and the loop linking β5 and β6 showed lower values, indicating a significant mobility in the ps-ns time scale, in agreement with the paucity of inter-residue NOEs in this region (Figure 1D). The single domains are arranged in a 'face to side' manner, similarly to what observed in the homologue PHD_VC5HCH_{NSD3} structure (65% sequence identity) (Supplementary Figures S1 and S3A). The fold similarity between the two NSD tandem domains is high, with a r.m.s.d. of 1.6 Å over 91 Cα atoms; major differences between the two structures are observable in the L3 loop, due to the presence of two consecutive Prolines (P1355-$P1356_{NSD3}$) in PHD_VC5HCH_{NSD3} occupied by an Arginine and a Proline in the equivalent positions in PHD_VC5HCH_{NSD1} (R2152-$P2153_{NSD1}$) (Supplementary Figure S3A). The 'face to side' orientation appears to be a hallmark of the NSD family, and differs from the 'face to back' arrangement adopted by the tandem PHD fingers of the histone acetyltransferase MOZ (monocytic leukemia Zinc-finger protein) (57,58) or of the adaptor component of BAF chromatin remodeling complex Dpf3b (59) (Supplementary Fig. S3B and C), thus defining a different structural class of PHD tandem domains.

Structural analysis of Sotos syndrome point mutations

Of the 17 Sotos syndrome missense mutations targeting PHD_VC5HCH_{NSD1}, 14 involve Zn^{2+}-coordinating residues and 3 affect amino-acids outside the metal binding site (Supplementary Table S1). As already observed in other PHD finger domains (30,31), the pathological effect of mutations involving Zn^{2+} coordinating Cysteines is ascribable to loss of function effects, due to domain unfolding caused by Zn^{2+} depletion. Conversely, the structural effect of Histidine mutations is less predictable, as they are often dispensable to preserve Zn^{2+} coordination and fold integrity (60) (see also further). In order to have a complete overview of the structural impact of the non-Cysteine pathological mutations we expressed and purified seven non-Cysteine Sotos mutants and compared their NMR spectra to the wild-type one (Figure 2A and B, Supplementary Figure S3E). The 1D-^{1}H spectra of $H2143E_{NSD1}$, $H2143Y_{NSD1}$, $H2162R_{NSD1}$ were characterized by distinctive line-broadening effects, indicative of aggregation and/or unfolding, thus revealing their fundamental role in Zn^{2+} coordination and domain stability (Figure 2A). Unexpectedly, $H2205R_{NSD1}$ mutation did not affect the domain fold as shown by the good peak dispersion of its 2D ^{1}H–^{15}N HSQC spectrum (Figure 2C). Conceivably, the three Zn^{2+} binding Cysteines are sufficient to maintain metal ion coordination, and the methylene groups of the Arginine-mutant can partially replace the hydrophobic interactions established by $H2205_{NSD1}$ with $M2177_{NSD1}$ (Figure 2B). Finally, among the mutants outside the Zn^{2+} binding sites, $Y2142N_{NSD1}$ and $F2182I_{NSD1}$ involved two conserved aromatic residues establishing stabilizing hydrophobic interactions with $F2122_{NSD1}$, $K2140_{NSD1}$, $W2157_{NSD1}$, $P2160_{NSD1}$, $W2161_{NSD1}$, $V2166_{NSD1}$, $M2177_{NSD1}$, $P2179_{NSD1}$, $M2190_{NSD1}$ and $L2191_{NSD1}$) (Figure 2B). Mutations of both $Y2142_{NSD1}$ and $F2182_{NSD1}$ resulted in domain unfolding and/or aggregation, as assessed by the broad lines and reduced peak dispersion of their ^{1}H-1D spectra (Figure 2A). Conversely, mutation of the solvent exposed $R2152_{NSD1}$ into Glutamine (Figure 2B, C), did not affect the domain fold, as assessed by the remarkable similarity between the wild-type and mutant ^{1}H–^{15}N-HSQC spectra (Figure 2C). In the context of Sotos syndrome, fold-destroying mutants most likely result in functional inactivation of NSD1, well in keeping with NSD1 haploinsufficiency being the primary pathogenic mechanism in the syndrome (61). However, a more complex scenario at molecular level could arise from fold-preserving mutations, whose effects might not be simply ascribed to a generic loss of function event. It is conceivable that the change in electrostatic potential induced by both the $H2205R_{NSD1}$ and $R2152Q_{NSD1}$ fold-preserving mutations results in downstream effects on the molecular complexes recruited by NSD1 and consequently on the transcriptional events regulated by these macromolecular assemblies. The functional impact of fold-preserving mutations has been already observed in other PHD finger containing proteins, such as Autoimmune regulator protein (AIRE), where PHD fingers pathological mutations acted on AIRE chromatin-associated interactome, thus reducing the activation of its target genes (31).

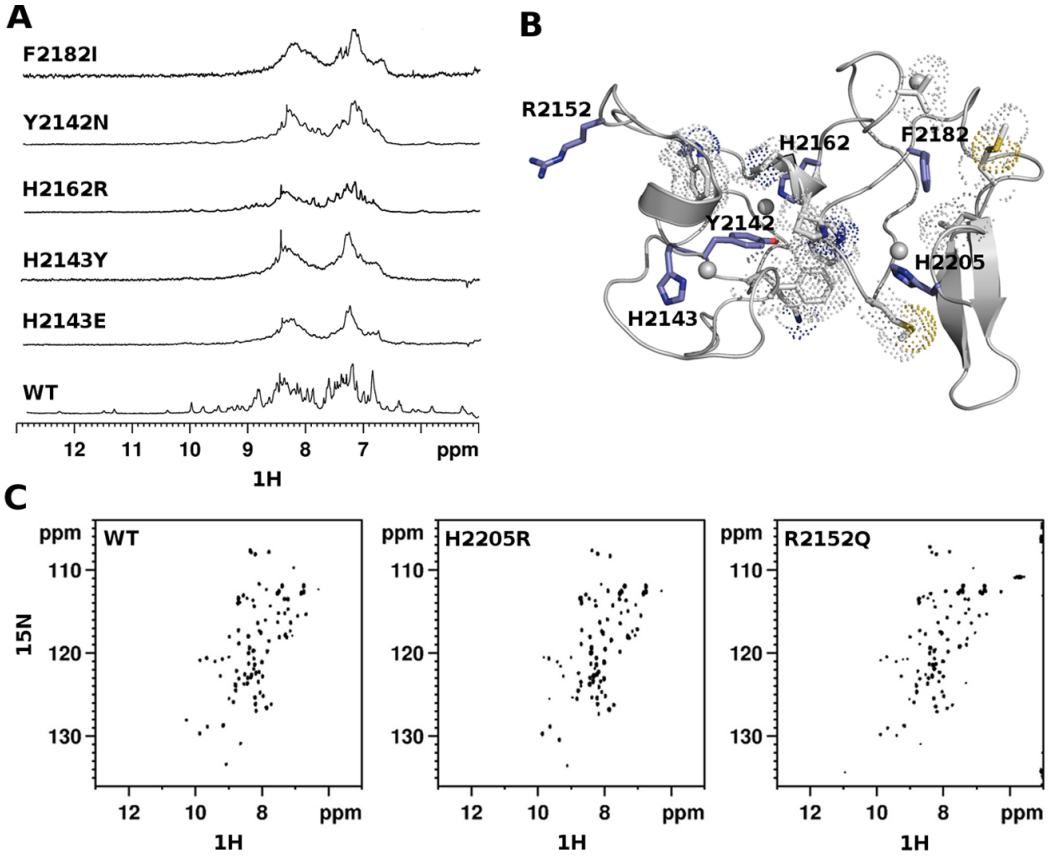

Figure 2. Structural effects of Sotos syndrome mutations. (**A**) 1D ^{1}H spectra (amide region) of wild-type and pathological mutants of PHD_VC5HCH_{NSD1}. (**B**) Cartoon representation of PHD_VC5HCH_{NSD1} (gray). The side chains of pathological point mutations are shown in blue sticks, residues involved in hydrophobic contacts are shown in grey sticks with dotted space-filled representation. (**C**) 2D ^{1}H–^{15}N HSQC spectra (amide region) of wild-type and pathological mutants of PHD_VC5HCH_{NSD1}.

PHD_VC5HCH_{NSD1} binds to $C2HR_{Nizp1}$ with low-micromolar affinity

Over the last decade PHD fingers as single modules or in combination with other chromatin binding domains have been intensively studied as histone binding recognition domains (62,63). Conversely, the possibility that PHD fingers can directly interact with non-histone proteins has been limited to only few examples, including the PHD fingers of MLL1 (28), of Pygo2 (27) and of Sp140 (64), interacting respectively with Cyp33, Bcl9 and Pin1. In line with the emerging multifaceted role of PHD fingers as structural hubs for multiple interactions, early yeast-two-hybrid screening and GST-pulldown assays, corroborated by site directed mutagenesis, have suggested a direct interaction between murine NSD1 and the nuclear repressor Nizp1 (also known as Zfp496) *via* the PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ domains (7,25). Prompted by this observation, we wondered whether we could confirm and thermodynamically characterize this PHD-finger/non-histone protein interaction. Indeed, evaluation of the binding through isothermal titration calorimetry indicated that a Nizp1 fragment (residues E397-$K434_{Nizp1}$ as defined in (7)) containing the putative Zn^{2+}-binding $C2HR_{Nizp1}$ domain binds to PHD_VC5HCH_{NSD1} with a dissociation constant of 3.8 ± 0.7 μM in a 1:1 ratio (Figure 3A). At room temperature ($T = 23°C$) the reaction is endothermic ($\Delta H = 4.1 \pm 0.1$ kcal/mol) and entropy driven ($T\Delta S = 11.4$ kcal/mol) suggesting a prominent role of hydrophobic interactions in complex formation.

Solution structure of $C2HR_{Nizp1}$

To get further molecular insights into the interaction between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ we first solved $C2HR_{Nizp1}$ solution structure by classical homonuclear and heteronuclear experiments (Table 2). Despite the presence of an Arginine ($R427_{Nizp1}$) replacing the fourth core Histidine in this atypical Zinc-finger sequence (Supplementary Figure S1B), the domain has a Zn^{2+} dependent fold, as assessed by the increased chemical shift dispersion of its 1D-^{1}H spectrum upon addition of one Zn^{2+} equivalent (Figure 4A). The domain adopts the typical Zinc-finger topology composed by a small two-stranded antiparallel β-sheet, encompassing residues Y405-$C407_{Nizp1}$ (β1) and G411-$F414_{Nizp}$ (β2), followed by a short positively charged α-helix (R417-$H423_{Nizp1}$) (Figure 4B). The N-terminus (E397-$K403_{Nizp1}$) and the C-terminus (R425-$K434_{Nizp1}$) are extremely flexible, in line with their low heteronuclear NOE values (Figure 4C). A network of hydrophobic interactions involving $F414_{Nizp1}$, $F420_{Nizp1}$, $L424_{Nizp1}$ also contributes to domain stability (Figure 4B). Of note, $W416_{Nizp1}$ is lo-

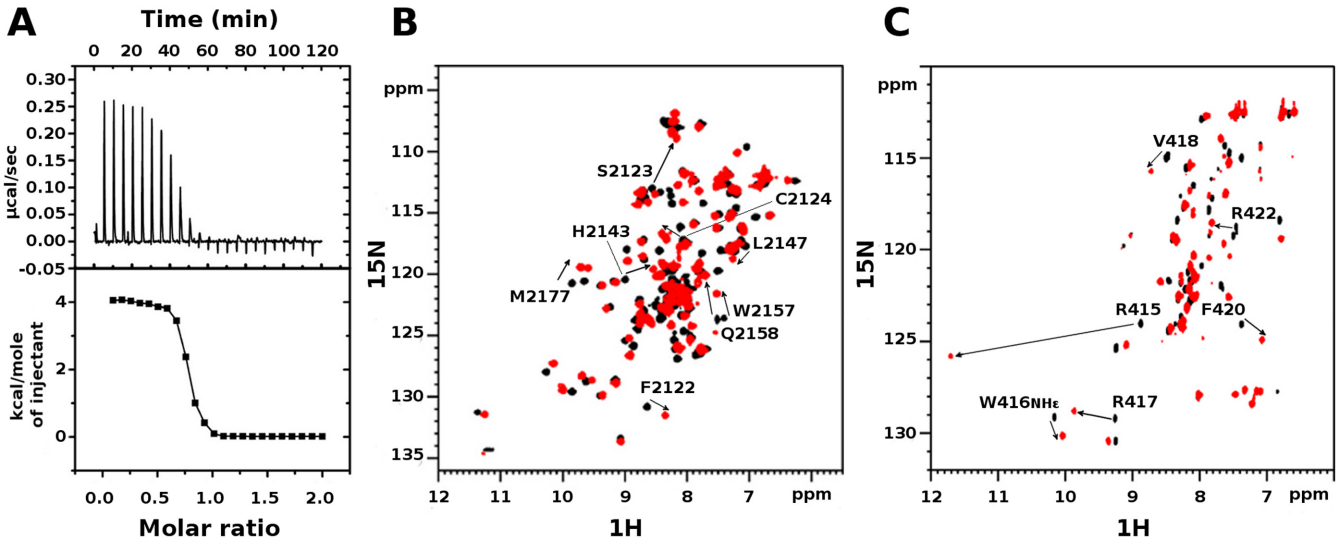

Figure 3. Interaction between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$. (**A**) ITC-binding curves of $C2HR_{Nizp1}$ to PHD_VC5HCH_{NSD1}. The upper panel shows the sequential heat pulses for domain–domain binding, and the lower panel shows the integrated data, corrected for heat of dilution and fit to a single-site-binding model using a nonlinear least-squares method (line). (**B**) Superposition of ^{1}H–^{15}N HSQC spectra of PHD_VC5HCH_{NSD1} without (black) and with (red) $C2HR_{Nizp1}$. (**C**) Superposition of ^{1}H-^{15}N HSQC spectra of $C2HR_{Nizp1}$ without (black) and with (red) PHD_VC5HCH_{NSD1}. The peak with the down shift to 11.8 ppm upon PHD_VC5HCH_{NSD1} addition corresponds to the backbone amide of $R415_{Nizp1}$, suggesting the presence of an H-bond upon complex formation. Peaks shifting more than <CSP>+1σ are explicitly labeled.

cated on a solvent exposed loop (RWR-loop) and protrudes out of the domain suggesting a possible functional role for this residue (Figure 4B). Finally, $C2HR_{Nizp1}$ is highly positively charged with a positive patch comprising the surface of the α-helix and the RWR-loop, well suited to interact with the negatively charged surface of PHD_VC5HCH_{NSD1} (Figure 4D). Of note, the RWR-loop appears to be well conserved along evolution, in line with its possible functional role (Supplementary Figure S1B). Preliminary structure calculations performed without including the Zn^{2+} ion revealed that three residues out of the $C2HR_{Nizp1}$ signature ($C407_{Nizp1}$, $C410_{Nizp1}$, $H423_{Nizp1}$) are near in space, at the appropriate distance to coordinate the metal ion, whereas $R427_{Nizp1}$ (Figure 4B), located on the flexible C-terminal tail is distant from these residues and is not involved in Zn^{2+} binding. We also excluded that $N409_{Nizp1}$ side chain (Figure 4B), which was in proximity to the metal binding site, could bind Zn^{2+} ion completing its coordination sphere, as its mutation into Alanine did not have any impact on the domain fold, as assessed by comparison between the mutant and wild-type 1D ^{1}H NMR spectra (Supplementary Figure S4A). Final structure calculations including distance restraints for the Zn-S^{-} and Zn-Nϵ, did not highlight any backbone carbonyl or suitable side chain in the proximity of the metal ion that could complete its coordination. Since the three-coordinated Zn^{2+} ion is partially solvent exposed, water molecules can easily fill its coordination sphere, as often observed in the catalytic Zn^{2+}-binding sites of enzymes or in other Zn^{2+} binding modules (60,65). Indeed, QM/MM (Quantum Mechanics/Molecular Mechanics) calculations confirmed the presence of one water molecule completing the tetrahedral Zn^{2+} ion coordination sphere (Figure 4B and Supplementary Figure S4B).

Table 2. Summary of conformational constraints and statistics for the 15 lowest energy structures of $C2HR_{Nizp1}$

Restraints Information [a,b]	<SA>[c]
Total number of experimental distance restraints	569
$C2HR_{Nizp1}$ (intraresidual/sequential/medium/long)	191/156/122/100
Zn^{2+} coordination restraints	3
Hydrogen bonds	1
Dihedral angle restraints (Φ/Ψ)	20/20
Deviation from idealized covalent geometry	
Bonds (Å)	0.0031 ± 0.0001
Angles (°)	0.481 ± 0.031
Coordinate R.m.s. Deviation (Å)[d]	
Ordered backbone atoms (N, Cα, C')	0.462 ± 0.146
Ordered heavy atoms	1.034 ± 0.34
Ramachandran Quality Parameters (%)[e]	
Residues in most favoured regions	93.00%
Residues in allowed regions	7.00%

[a] No dihedral angle restraints in any of the structures included in the ensemble was violated by more than 5°.
[b] Simulated annealing, statistics refers to the ensemble of 15 structures with the lowest energy.
[c] No distance restraint in any of the structures included in the ensemble was violated by more then 0.5 Å.
[d] Root mean squared deviation between the ensemble of structures <SA> and the lowest energy structure and Ramachandran quality parameters calculated on residues Ser404–Ser426.

Mapping of PHD_VC5HCH_{NSD1}/$C2HR_{Nizp1}$ interaction surface

To map the interaction surface between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ we performed CSP experiments titrating ^{15}N-PHD_VC5HCH_{NSD1} with unlabeled $C2HR_{Nizp1}$. The

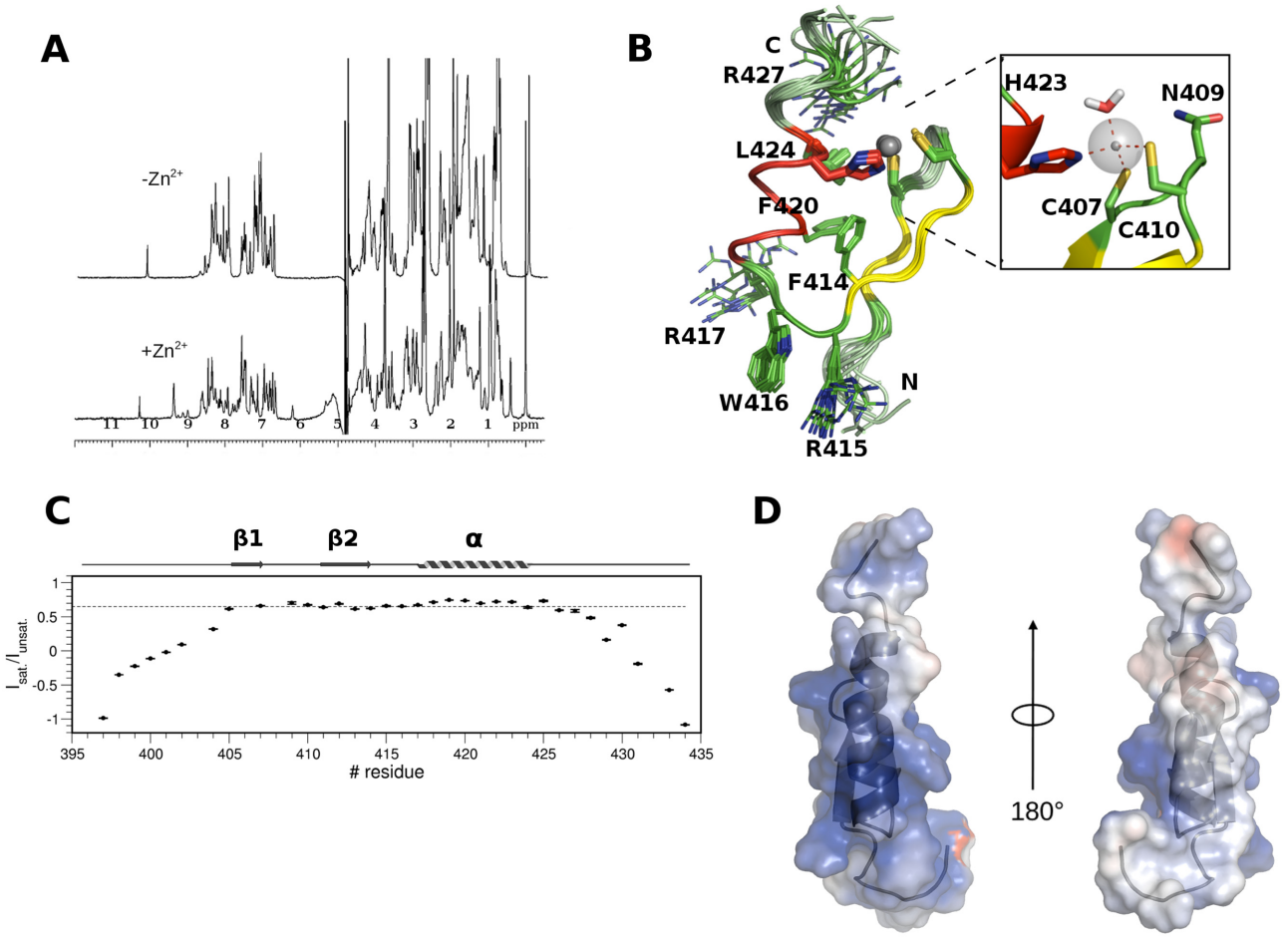

Figure 4. Solution structure of $C2HR_{Nizp1}$. (**A**) 1D ^{1}H spectrum of a synthetic peptide corresponding to $C2HR_{Nizp1}$ (0.2 mM) with (bottom) and without (top) one equivalent of $ZnCl_2$. (**B**) Superposition of the best 15 NMR structures of $C2HR_{Nizp1}$, with the two β-strands, the α helix and the loops coloured in yellow, red and green, respectively. Zn^{2+} binding residues and the Zn^{2+} ions are represented in sticks and spheres, respectively. The hydrophobic core, the RWR_{Nizp1} loop and $R427_{Nizp1}$ are represented in lines. A zoom into the Zn^{2+} binding site of a representative structure from QM/MM calculations is shown in the inset, highlighting the presence of one water molecule completing the tetrahedral Zn^{2+} ion coordination sphere (highlighted by red dotted lines). $H423_{Nizp1}$, $C407_{Nizp1}$, $N409_{Nizp1}$, $C410_{Nizp1}$ and the water molecule are represented in sticks. (**C**) Backbone dynamics of $C2HR_{Nizp1}$. The dotted line indicates the $\{^{1}H\}$–^{15}N heteronuclear NOE value threshold of 0.65. Elements of secondary structure are indicated on the top of the figure. (**D**) Electrostatic surface potential of $C2HR_{Nizp1}$.

CSP method is extremely sensitive and versatile, well suited to detect both strong and weak interactions, to map binding sites and detect residues both directly interacting with the ligand and/or indirectly affected by the association. Considerable changes in the slow- to intermediate-exchange regime occurred in PHD_VC5HCH_{NSD1} ^{1}H–^{15}N HSQC spectrum upon addition of unlabeled $C2HR_{Nizp1}$ (Figure 3B). At equimolar ratio PHD_VC5HCH_{NSD1} resonances with significant amide CSP or disappearing upon binding, clustered on the negatively charged surface of the tandem domain, mainly involving residues at the interface between PHD_V and C5HCH ($D2119_{NSD1}$, $F2122_{NSD1}$, $S2123_{NSD1}$, $C2124_{NSD1}$, $H2143_{NSD1}$, $L2147_{NSD1}$, $W2157_{NSD1}$, $E2158_{NSD1}$, $M2177_{NSD1}$, $E2204_{NSD1}$) (Figure 5A and B). The reverse NMR titration, in which unlabelled PHD_VC5HCH_{NSD1} was stepwise added into ^{15}N-$C2HR_{Nizp1}$ (Figure 3C) indicated that the $C2HR_{Nizp1}$ α-helix and the solvent exposed RWR-loop were affected by the binding, as assessed by the remarkable amide chemical shifts of residues $R415_{Nizp1}$, $W416_{Nizp1}$, $R417_{Nizp1}$, $V418_{Nizp1}$, $F420_{Nizp1}$, $R422_{Nizp1}$ (Figure 5C and D). Relaxation measurements acquired on ^{15}N-PHD_VC5HCH_{NSD1} or on ^{15}N-$C2HR_{Nizp1}$ in the presence of unlabelled $C2HR_{Nizp1}$ or PHD_VC5HCH_{NSD1}, respectively, yielded a correlation time of ∼12 ns, well in agreement with a 1:1 complex (MW ∼15 kDa) tumbling in solution as a unique structural entity (Supplementary Figure S2 A and B).

Three-dimensional model of PHD_VC5HCH_{NSD1}/$C2HR_{Nizp1}$ complex

Although most of the backbone resonances assignment of bound PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ was accomplished, due to unfavourable dynamic regime and to sample precipitation and degradation at concentrations beyond 0.3 mM, it was not possible to detect intermolecular NOEs to determine the complex structure using the standard NOE-based approach. Thus, we took advantage of

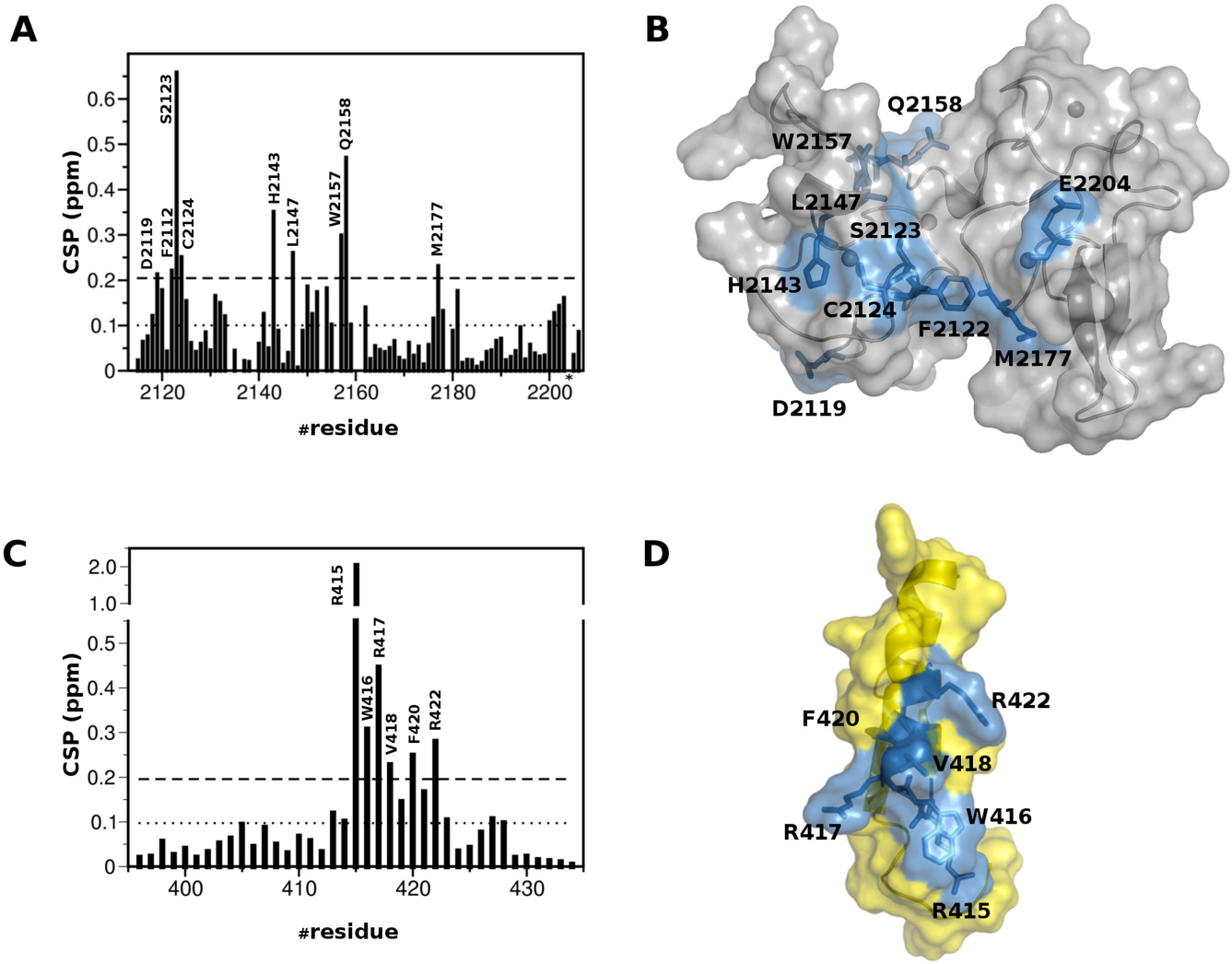

Figure 5. Mapping of the interaction between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$. (**A**) Histograms showing the averaged backbone chemical shift differences observed in PHD_VC5HCH_{NSD1} ^1H–^{15}N HSQC (0.2 mM) upon addition of a 2-fold excess of $C2HCR_{Nizp1}$. Missing bars correspond to either Prolines or amides that are not visible because of exchange with the solvent. The asterisk indicates peak disappearance ($Glu2204_{NSD1}$) upon binding. The last two bars in the plot correspond to the indole amides of $W2157_{NSD1}$ and $W2168_{NSD1}$, respectively. (**B**) Cartoon and surface representation of PHD_VC5HCH_{NSD1}. Residues showing significant chemical shift perturbation are represented in sticks and coloured in blue. (**C**) Histograms showing the averaged backbone chemical shift differences observed in $C2HCR_{Nizp1}$ ^1H–^{15}N HSQC (0.2mM) upon addition of a 2-fold excess of PHD_VC5HCH_{NSD1} In all the histograms, the dotted line represents <CSP>, the dashed line represents <CSP> + 1σ value, where σ corresponds to standard deviation. (**D**) Cartoon and surface representation of $C2HR_{Nizp1}$ (yellow). Residues showing significant chemical shift perturbation are represented in sticks and are coloured in blue.

the well-established HADDOCK 2.0 software (51) to obtain a structural model of the complex, exploiting experimental CSP data, as described in Materials and Methods. Docking models were clustered based on the interface r.m.s.d. resulting in 11 clusters (Figure 6A, Supplementary Table S2). Importantly, in all the clusters $C2HR_{Nizp1}$ docks inside PHD_VC5HCH_{NSD1} placing its α helix in the interdomain groove formed by PHD_V and C5HCH (Figure 6B). In particular in Cluster 1, showing the best HADDOCK score (Figure 6A, Supplementary Table S2), $C2HR_{Nizp1}$ interacts with PHD_VC5HCH_{NSD1} inserting its $W416_{Nizp1}$ side chain into the hydrophobic pocket generated by $H2205_{NSD1}$, $F2122_{NSD1}$, $M2177_{NSD1}$, $P2179_{NSD1}$, $R415_{Nizp1}$ and $V418_{Nizp1}$ (Figure 6C-left). A set of polar contacts involving the Arginines of the RWR signature seems to contribute to the binding: the guanidinium group of $R417_{Nizp1}$ forms a salt-bridge with the carboxylate of $E2204_{NSD1}$; moreover $R415_{Nizp1}$ establishes electrostatic interactions with the carboxylate of $E2120_{NSD1}$ through its guanidinium group and forms through its amide proton a hydrogen bond with the carbonyl of $S2123_{NSD1}$ (Figure 6C, middle). Finally, $R422_{Nizp1}$ points toward the interdomain groove, interacting with $L2147_{NSD1}$ and $W2157_{NSD1}$ (Figure 6C right). All the residues establishing favourable intermolecular interactions were consistent with the observed chemical shift perturbations. In an effort to assess the validity of our HADDOCK model we made a series of single point mutations around the binding interface of both PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ to investigate their effect in NMR and/or calorimetric titrations (Table 3). In

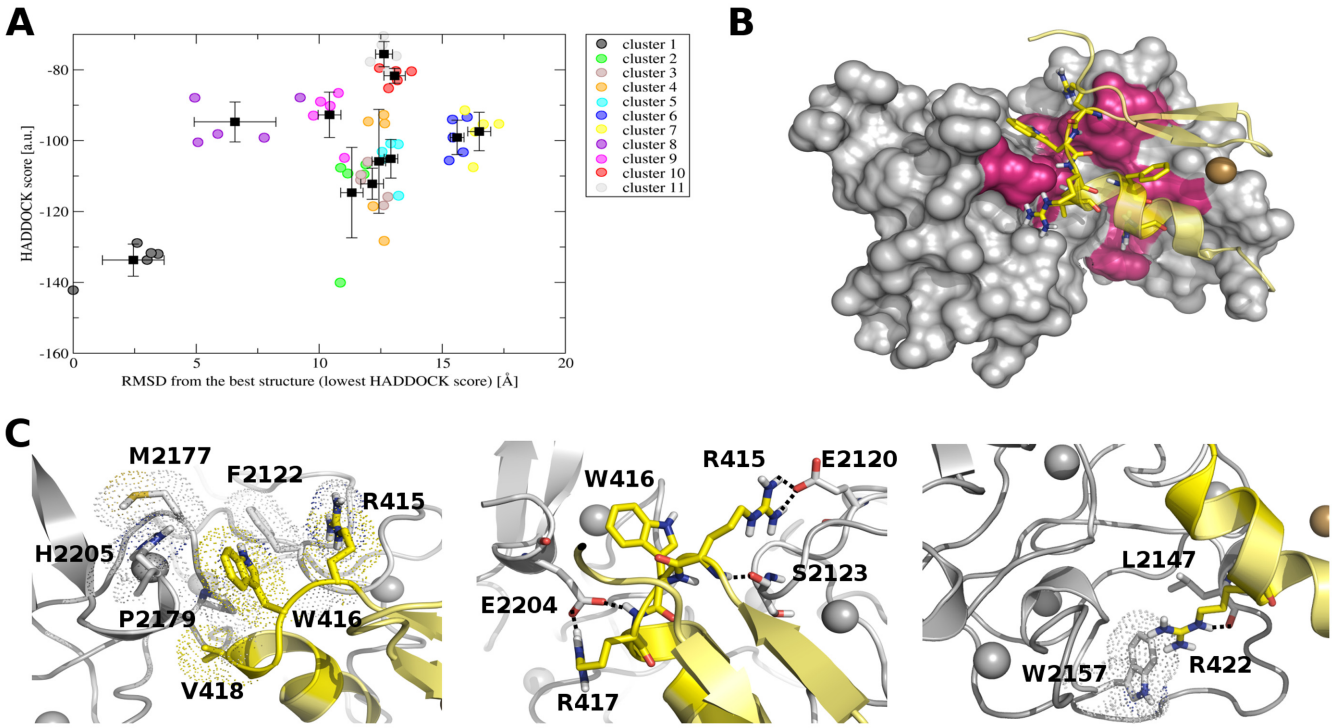

Figure 6. HADDOCK model of PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$. (**A**) HADDOCK score versus r.m.s.d. from the lowest energy complex structure in terms of HADDOCK score (a.u.). Circles correspond to the best five structures of each cluster, black squares correspond to the cluster averages with the standard deviation indicated by bars. (**B**) Representative pose of the best docking model in terms of HADDOCK score. PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ are shown as surface and cartoon, respectively. Active residues are shown in hot pink and yellow sticks in PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$, respectively. (**C**) Details of the relevant interactions between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$. Hydrophobic and electrostatic interactions between RWR_{Nizp1} loop and PHD_vC5HCH_{NSD1} are shown on the left and central panels, respectively. Interactions of $R422_{Nizp1}$ are shown on the right panel. Hydrophobic interactions are highlighted with dotted space-filling representation and polar interactions with dotted lines. Zn^{2+} ions are shown as grey ($PHD_VC5HCH_{NSD1)}$ and gold spheres ($C2HR_{Nizp1}$)

particular, mutation of the exposed $W416_{Nizp1}$ into Alanine had a very strong impact, with little or virtually no detectable binding in NMR titrations, implying almost complete abrogation of the interaction (Supplementary Figure S5B, Table 3). This result strongly supports the role of hydrophobic interaction in this binding process, in agreement with the entropically driven reaction, most likely deriving from the desolvation of the intermolecular binding interface. Both $R417_{Nizp1}$ and $E2204_{NSD1}$ mutations into Alanine resulted in one order of magnitude reduction in binding affinity and a reduction of the average chemical shifts displacements (Table 3) corroborating the presence of a salt-bridge between these two residues (Figure 6C, left). Mutation of $R415_{Nizp1}$ did not have a relevant impact in binding, suggesting that the major interaction with PHD_VC5HCH_{NSD1} occurs via a hydrogen bond between its amide and the carbonyl of $S2123_{NSD1}$, in line with the extreme down-shift of $S2123_{NSD1}$ amide proton resonance upon binding (Figures 3C and 6C, middle). $R422A_{Nizp1}$, with only a 2-fold increase in the Kd, had a minor impact in binding (Table 3), implying a secondary role for this residue in substrate recognition. Unfortunately, we could not assess the role of $E2120_{NSD1}$, $F2122_{NSD1}$ and $S2123_{NSD1}$ in binding as their mutations into Alanine caused low expression

Table 3. Thermodynamic parameters, stoichiometry (n) and dissociation constants (Kd) between PHD_VC5HCH_{NSD1} and $C2HR_{Nizp1}$ wild type (WT) and mutants measured by ITC ($T = 22°C$)

PHD_VC5HCH_{NSD1}	$C2HR_{Nizp1}$	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG (kcal/mol)	n	K_d (μM)
WT	WT	4.1 ± 0.1	11.4± 0.1	-7.3 ± 0.1	1.1 ± 0.1	3.8 ± 0.7
R2152Q	WT	3.5 ± 0.2	10.7 ± 0.1	-7.2 ± 0.1	1.0 ± 0.1	4.9 ± 0.6
E2204A	WT	0.4 ± 0.1	6.3 ± 0.2	-5.9 ± 0.2	0.9 ± 0.1	45.2 ± 6.9
H2205R	WT	1.0 ± 0.1	7.2 ± 0.1	6.2 ± 0.1	1.0 ± 0.1	27.2 ± 5.3
WT	R415A	2.2 ± 0.1	9.3 ± 0.1	-7.1 ± 0.1	1.1 ± 0.1	6.1 ± 0.4
WT	W416A	N.D.	N.D.	N.D.	N.D.	N.D.
WT	R417A	1.5 ± 0.1	7.5 ± 0.1	6.0 ± 0.1	1.1 ± 0.1	36.0 ± 6.0
WT	R422A	0.7 ± 0.1	7.7 ± 0.1	7.0 ± 0.1	0.9 ± 0.1	7.0 ± 0.2
WT	R427A	1.7 ± 0.1	9.0 ± 0.1	-7.3 ± 0.1	1.0 ± 0.1	4.0 ± 0.3

yields and partial PHD_VC5HCH_{NSD1} unfolding. Mutation of the $R427_{Nizp1}$ within the C2HR signature did not influence binding, in agreement with both the model and the chemical shift mapping which did not highlight any relevant role of this residue in the binding. We also verified the impact of fold-preserving Sotos syndrome point mutations in the binding to $C2HR_{Nizp1}$. As expected, $R2152Q_{NSD1}$ mutation, located on the opposite surface with respect to $C2HR_{Nizp1}$ binding site, did not affect complex formation (Table 3). Conversely, $H2205R_{NSD1}$ mutation, located near the binding site identified by CSP, reduced the binding of almost one order of magnitude, probably because of repulsive interactions with $R417_{Nizp1}$ and of the loss of stabilizing hydrophobic interactions between $H2205_{NSD1}$ and $W416_{Nizp1}$ (Figure 6C, left). Based on these findings and on the repressive role attributed to Nizp1 and NSD1 interaction (7), it is conceivable that Sotos syndrome mutations targeting PHD_VC5HCH_{NSD1} and hampering interaction with Nizp1 might have an impact on the repression of growth-promoting genes, eventually resulting in overgrowth disease.

Biophysical assessment of PHD_VC5HCH_{NSD1} interaction with histone H3 N-terminal tail peptides

Until now biochemical assays aiming at the *in vitro* assessment of the effective histone binding ability of PHD_VC5HCH_{NSD1} were not conclusive. On one hand recent pulldown assays of GST-PHD_VC5HCH_{NSD1} with unfractionated calf thymus histones or with biotinylated histone H3 peptides did not provide evidence of histone binding (26). On the other hand histone binding assays using MODified Histone Peptide Array TM (Activ Motif) (Supplementary Figure S6A,B) were in line with previous pull-down assays of GST-PHD_VC5HCH_{NSD1} with modified biotinylated peptides (21 aminoacids long) (25), suggesting an interaction preference for H3K4me3 and H3K9me3. Because of the qualitative nature of biochemical binding assays, that are often prone both to false positive and negative results (66–68), we biophysically verified this putative interaction using NMR CSP. At variance to what observed in titrations with $C2HR_{Nizp1}$, addition of up to 12 molar excess of unmodified or modified histone peptides ($H3_{1–21}$, $H3K4me3_{1–21}$ and $H3K9me3_{1–21}$) induced in PHD_VC5HCH_{NSD1} $^1H–^{15}N$ HSQC spectrum very small changes in the fast exchange regime, suggestive of extremely weak interactions (with an average CSP value <CSP> of ∼0.02 ppm) (Figure 7A, Supplementary Figure S6C). As the binding curves did not reach saturation, it was not possible to reliably fit the binding curves obtained during the individual titrations, implying affinities in the high mM range, i.e. orders of magnitude higher with respect to the micromolar affinity usually observed in PHD-histone tail interactions (69,70). Accordingly, isothermal titration calorimetry experiments did not develop sufficient heat to allow Kd determination (data not shown). Notably, in all the titrations, irrespectively of the peptides modification status, residues exhibiting significant amide chemical shift (CSP > <CSP> + 1σ) did not involve the classical histone binding surface, composed by the N-terminus of the PHD finger, the first β-strand, and the conserved PXGXW binding pocket. Peaks significantly shifting upon H3 peptides addition belonged to residues located on both PHD_V and C5HCH ($D2119_{NSD1}$, $E2120_{NSD1}$, $S2123_{NSD1}$, $C2124_{NSD1}$, $T2150_{NSD1}$, $K2151_{NSD1}$, $W2157_{NSD1}$, $E2158_{NSD1}$, $H2162_{NSD1}$, $C2164_{NSD1}$, $D2165_{NSD1}$, $S2181_{NSD1}$, $H2205_{NSD1}$), mainly, at the interface between PHD_V and C5HCH, on the negatively charged surface of the tandem domain (Figures 1C and 7B). Of note, these residues did not form a continuous interaction surface when mapped on the surface of the tandem domain (Figure 7B). Intriguingly, NMR titrations performed reducing or increasing the number of positively charged residues using shorter ($H3_{1–10}$) or longer peptides ($H3_{1–36}$), did not change the chemical shift differences profiles, but rather resulted in smaller and higher chemical shifts displacements, respectively (Supplementary Figure S6D). Taken together, these observations strongly support the non-specific electrostatic nature of PHD_VC5HCH_{NSD1} histone H3 tail interaction. Indeed, titration of the domain with up to 300 mM of Lysine induced amide chemical shifts in the same region (Supplementary Figure S6D). Overall, in contrast to what observed in previous binding assays with biotinylated histone peptides (25), but in agreement with recent GST-pulldown experiments with unfractionated calf thymus histones and biotinylated histone peptides (26), our biophysical data exclude a role of PHD_VC5HCH_{NSD1} in decoding methylated/unmethylated H3K4 and H3K9 epigenetic marks. We cannot rule out that, despite the extensive modification coverage guaranteed by the peptides arrays, PHD_VC5HCH_{NSD1} has an as-of-yet unidentified histone modification(s) specificity. It is conceivable that the specific recruitment of full-length NSD1 to chromatin might be mediated by the combinatorial interactions of the other chromatin readers (four PHD fingers and two PWWP) present in the full-length protein. Herein, the acidic surface of PHD_VC5HCH_{NSD1} might work as an electrostatic platform reinforcing the interaction with the basic histone tails. Indeed, it is not unusual that isolated readers displaying *in vitro* weak or non-specific interactions with their histone peptide target, in the context of the full length protein reach physiologically relevant binding affinities range, thanks to multivalent interactions and/or contacts using scaffolding domains that bridge their host proteins with other subunits of the chromatin bound complex (62). As the $C2HR_{Nizp1}$ binding site involves PHD_VC5HCH_{NSD1} interdomain space, we wondered whether the weak interaction with the histone H3 N-terminal tail could compete with complex formation. The $^1H–^{15}N$ HSQC spectra of both ^{15}N-PHD_VC5HCH_{NSD1}/$C2HR_{Nizp1}$, and of ^{15}N-$C2HR_{Nizp1}$/PHD_VC5HCH_{NSD1} (1:1) did not show any substantial change upon addition of an excess of $H3_{1–21}$ peptide (1:12), excluding a competing influence of the histone H3 tail on PHD_VC5HCH_{NSD1}/$C2HR_{Nizp1}$ interaction (Supplementary Figure S5A).

Functional divergence of the PHD_VC5HCH domain within the NSD family

Despite the high sequence similarity shared by the PHD_VC5HCH domain within the NSD family (PHD_VC5HCH_{NSD1} shares 61% and 64% identity with PHD_VC5HCH_{NSD2} and PHD_VC5HCH_{NSD3}, re-

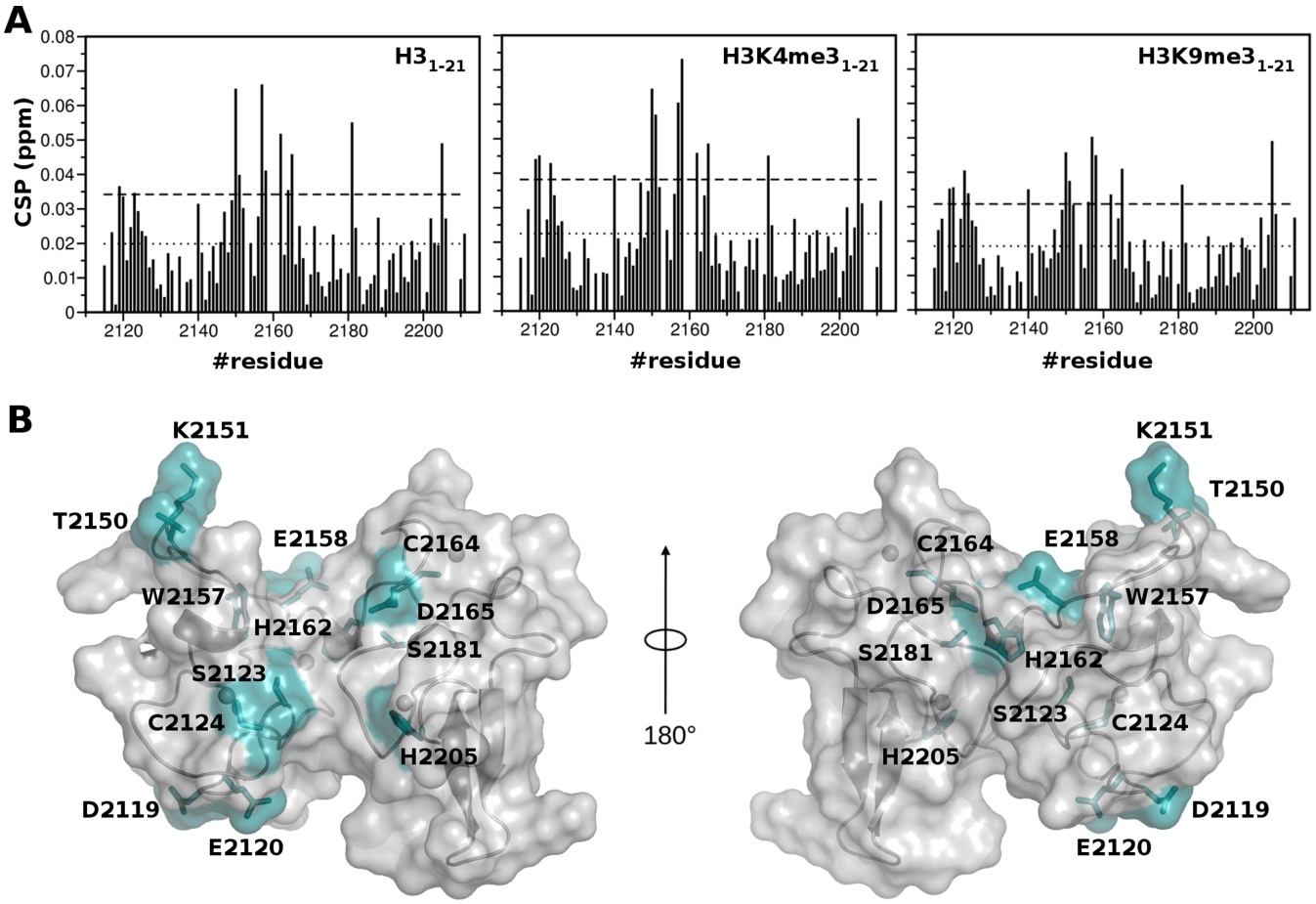

Figure 7. Mapping of the interaction with histone H3 amino-terminal peptides. (**A**) Histograms showing the averaged backbone chemical shift perturbations (CSP) observed in ^{1}H–^{15}N HSQC of PHD_VC5HCH_{NSD1} (0.2 mM) upon addition of a twelve-fold excess of $H3_{1–21}$, $H3K4me3_{1–21}$, $H3K9me3_{1–21}$. The last two bars in the plot correspond to the indole amides of $W2157_{NSD1}$ and $W2168_{NSD1}$, respectively. The dotted line represents <CSP>, the dashed line represents <CSP> + 1σ, where σ corresponds to standard deviation. Missing bars correspond to either Prolines or amides that are not visible because of exchange with the solvent. (**B**) Surface and cartoon representation of PHD_VC5HCH_{NSD1}. Residues showing significant chemical shift perturbation (CSP > <CSP>+1σ) are represented in cyan.

spectively) (Supplementary Figure S1A), they appear to have evolved to exert different functions. As a matter of fact, PHD_VC5HCH_{NSD1} does not function as classical epigenetic reader, whereas PHD_VC5HCH_{NSD2} and PHD_VC5HCH_{NSD3} preferentially recognize H3K4me0 (demonstrated only biochemically) and H3K9me3 marks, respectively (26). Based on sequence comparison and on the available structural knowledge (this study and the 3D structure of the PHD_VC5HCH_{NSD3}/H3K9me3 complex (26)), we hypothesized that PHD_VC5HCH_{NSD1} inability to efficiently recognize the positively charged histone H3 tail could be ascribable to electrostatic repulsions, generated by two basic residues ($R2117_{NSD1}$, $K2134_{NSD1}$) on the canonical histone interaction surface (Supplementary Figures S1A, S3D). Notably, the equivalent positions in NSD2 and NSD3 are occupied by a neutral aminoacid and by an acidic residue forming a salt-bridge with H3R2, as highlighted by the PHD_VC5HCH_{NSD3}/H3K9me3 crystallographic structure (Supplementary Figure S3D). However, simultaneous mutations of these basic residues ($R2117A_{NSD1}$, $K2134D_{NSD1}$) aiming at converting PHD_VC5HCH_{NSD1} into a classical histone H3 reader, failed to induce the expected functional twist (Supplementary Figure S6E). Collectively, these results suggest that a more complex combination of amino-acids substitutions, possibly related to yet unidentified conformational and dynamic effects might dictate the different histone recognition propensities within the NSD family. Following the same line, PHD_VC5HCH domain functional divergence seems also to reflect into different $C2HR_{Nizp1}$ binding abilities attributed to the three NSD proteins. In fact yeast two hybrid screenings on both $C5HCH_{NSD2}$ and $C5HCH_{NSD3}$ (71) along with GST-pulldown assays on PHD_VC5HCH_{NSD3} (26) did not support their direct interaction with $C2HR_{Nizp1}$. Conceivably, also in this case a combination of dynamical effects and of minor sequence variations targeting the interaction surface might account for possible differences in $C2HR_{Nizp1}$ binding specificity within the NSD family. Future structural, dynamical and biophysical studies i. investigating the effective binding ability of the other NSD members to $C2HR_{Nizp1}$ and ii. verifying the putative histone H3 interaction with PHD_VC5HCH_{NSD2} are warranted to complete our picture on the structure-function relationship within the PHD_VC5HCH-NSD family.

Concluding remarks

In the last years, PHD fingers, alone or in tandem with other chromatin binding domains, have emerged as multifaceted interaction platforms, whose functions go far beyond the perceiving of the epigenetic landscape (62,63). As a matter of fact PHD fingers have been shown to work often as scaffolding domains bridging their host proteins with other subunits of bigger macromolecular complexes (62,63,72). Herein, PHD_VC5HCH_{NSD1} seems to represent a paradigmatic example for the PHD domain functional versatility: on one hand it has been proposed to decode the methylation status of histone H3K4 and H3K9 (25), on the other hand it has been shown to mediate the recruitment of NSD1 to reinforce Nizp1 transcriptional repressor activity via its direct binding to $C2HR_{Nizp1}$ Zinc-finger domain (25,29,73). Notably, structural and biophysical studies proving all these interactions and providing molecular details on them were lacking. Here, we have shown that the tandem domain PHD_VC5HCH_{NSD1} forms an indivisible structural entity that interacts only very weakly with histone H3 N-terminal tail peptides. The binding occurs via non-specific electrostatic forces, irrespectively of histone H3 methylation and involves a distinct surface with respect to the classical histone H3 tail binding groove. While the role of PHD_VC5HCH_{NSD1} as classical epigenetic reader is highly questionable, ITC and NMR CSP experiments clearly demonstrate a finger-finger micromolar interaction between $C2HR_{Nizp1}$ and PHD_VC5HCH_{NSD1}. Via a data driven docking model, supported by site-directed mutagenesis, we have provided first molecular insights into the recruitment of the co-repressor NSD1 to the repressor Nizp1 via PHD_VC5HCH_{NSD1} and a solvent exposed RWR motif of $C2HR_{Nizp1}$. In particular, the RWR loop of Nizp1 accommodate into the PHD_VC5HCH_{NSD1} interdomain groove establishing hydrophobic and electrostatic interactions. Interestingly, both $C5HCH_{NSD1}$ and $C2HR_{Nizp1}$ domains have been acquired late in evolution (7) and their high conservation in vertebrates support an evolutionary preservation of this protein-protein interaction (Supplementary Figure S1). Recruitment of co-repressors through short sequence motifs is a recurrent theme in transcriptional repression, as observed for example in the engagement of Groucho corepressor via the WPRW motif of Drosophila transcriptional repressors Dorsal and Hairy (74). The prominent functional role exerted by the RWR_{Nizp1} signature within the $C2HR_{Nizp1}$ domain strengthens also the notion that this atypical Zinc-finger motif is more than just a degenerate evolutionary vestige of the classical Zinc-finger C2H2 motif (7). This is in keeping with previous studies showing that Nizp1 transcriptional repressor activity depends on C2HR domain integrity, as fold destroying mutations abolished its binding to NSD1, thus reducing transcriptional repression (7). Importantly, in the Sotos syndrome context we have shown that the fold preserving mutation $H2205R_{NSD1}$ also diminishes the interaction with $C2HR_{Nizp1}$ supporting a mechanistic scenario in which impairment of the repressive NSD1/Nizp1 complex might alter the physiological down-regulation of growth-promoting genes, thus resulting in the overgrowth phenotype, typical of Sotos syndrome. In this context, it is conceivable that other nuclear proteins besides Nizp1 bind to NSD1, whereby the PHD finger domains, in line with their functional versatility, might mediate additional protein interactions. Future studies dedicated to NSD1/Nizp1 dependent transcriptome and interactome in normal and pathological conditions will be instrumental in understanding the patho-physiological role of NSD1/Nizp1 interaction and in devising new strategies for successful treatments in overgrowth diseases.

ACCESSION NUMBERS. DB accession codes: 2naa; 2nab BioMagResBank 25933; 25934.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We wish to thank the CERM Infrastructure (Florence) for access to NMR instrumentation. D.S. conducted this study as partial fulfillment of his PhD in Molecular Medicine, Program in Cellular and Molecular Biology, Vita e Salute San Raffaele University, Milan, Italy.

FUNDING

Italian Ministry for Health [RF-2013–02354880 to G.M and G.T.]; Italian Association Cancer Research [AIRC-13149 to G.M.]; Fondazione Telethon [TCP99035 to G.M.]; Fondazione Veronesi [to F.M.G.]; Marie Curie Co-funding of Regional, National and International Programmes [PCOFUND-GA-2010-267264 INVEST to M.A.C.R.]; Fondazione Cariplo [2009-3577 to M.G.]. Funding for open access charge: Italian Ministry for Health [RF-2013-02354880] or Italian Association Cancer Research [AIRC-13149].

Conflict of interest statement. None declared.

REFERENCES

1. Wagner,E.J. and Carpenter,P.B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.*, **13**, 115–126.
2. Vougiouklakis,T., Hamamoto,R., Nakamura,Y. and Saloura,V. (2015) The NSD family of protein methyltransferases in human cancer. *Epigenomics*, **7**, 863–874.
3. Morishita,M. and di Luccio,E. (2011) Cancers and the NSD family of histone lysine methyltransferases. *Biochim. Biophys. Acta*, **1816**, 158–163.
4. Kurotaki,N., Harada,N., Yoshiura,K., Sugano,S., Niikawa,N. and Matsumoto,N. (2001) Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. *Gene*, **279**, 197–204.
5. Huang,N., vom Baur,E., Garnier,J.M., Lerouge,T., Vonesch,J.L., Lutz,Y., Chambon,P. and Losson,R. (1998) Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.*, **17**, 3398–3412.
6. Rayasam,G.V., Wendling,O., Angrand,P.O., Mark,M., Niederreither,K., Song,L., Lerouge,T., Hager,G.L., Chambon,P. and Losson,R. (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.*, **22**, 3153–3163.
7. Nielsen,A.L., Jorgensen,P., Lerouge,T., Cervino,M., Chambon,P. and Losson,R. (2004) Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. *Mol. Cell. Biol.*, **24**, 5184–5196.

8. Berdasco,M., Ropero,S., Setien,F., Fraga,M.F., Lapunzina,P., Losson,R., Alaminos,M., Cheung,N.K., Rahman,N. and Esteller,M. (2009) Epigenetic inactivation of the sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 21830–21835.
9. Lucio-Eterovic,A.K., Singh,M.M., Gardner,J.E., Veerappan,C.S., Rice,J.C. and Carpenter,P.B. (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 16952–16957.
10. Yang,H., Pesavento,J.J., Starnes,T.W., Cryderman,D.E., Wallrath,L.L., Kelleher,N.L. and Mizzen,C.A. (2008) Preferential dimethylation of histone H4 lysine 20 by Suv4–20. *J. Biol. Chem.*, **283**, 12085–12092.
11. Morishita,M., Mevius,D. and di Luccio,E. (2014) In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L. *BMC Struct. Biol.*, **14**, 25.
12. Morishita,M. and di Luccio,E. (2011) Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. *Biochem. Biophys. Res. Commun.*, **412**, 214–219.
13. Qiao,Q., Li,Y., Chen,Z., Wang,M., Reinberg,D. and Xu,R.M. (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.*, **286**, 8361–8368.
14. Jaju,R.J., Fidler,C., Haas,O.A., Strickson,A.J., Watkins,F., Clark,K., Cross,N.C., Cheng,J.F., Aplan,P.D., Kearney,L. *et al.* (2001) A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood*, **98**, 1264–1267.
15. Hollink,I.H., van den Heuvel-Eibrink,M.M., Arentsen-Peters,S.T., Pratcorona,M., Abbas,S., Kuipers,J.E., van Galen,J.F., Beverloo,H.B., Sonneveld,E., Kaspers,G.J. *et al.* (2011) NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood*, **118**, 3645–3656.
16. Thol,F., Kolking,B., Hollink,I.H., Damm,F., van den Heuvel-Eibrink,M.M., Michel Zwaan,C., Bug,G., Ottmann,O., Wagner,K., Morgan,M. *et al.* (2013) Analysis of NUP98/NSD1 translocations in adult AML and MDS patients. *Leukemia*, **27**, 750–754.
17. Kurotaki,N., Harada,N., Yoshiura,K., Sugano,S., Niikawa,N. and Matsumoto,N. (2001) Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. *Gene*, **279**, 197–204.
18. Kurotaki,N., Imaizumi,K., Harada,N., Masuno,M., Kondoh,T., Nagai,T., Ohashi,H., Naritomi,K., Tsukahara,M., Makita,Y. *et al.* (2002) Haploinsufficiency of NSD1 causes sotos syndrome. *Nat. Genet.*, **30**, 365–366.
19. Nagai,T., Matsumoto,N., Kurotaki,N., Harada,N., Niikawa,N., Ogata,T., Imaizumi,K., Kurosawa,K., Kondoh,T., Ohashi,H. *et al.* (2003) Sotos syndrome and haploinsufficiency of NSD1: Clinical features of intragenic mutations and submicroscopic deletions. *J. Med. Genet.*, **40**, 285–289.
20. Douglas,J., Hanks,S., Temple,I.K., Davies,S., Murray,A., Upadhyaya,M., Tomkins,S., Hughes,H.E., Cole,T.R. and Rahman,N. (2003) NSD1 mutations are the major cause of sotos syndrome and occur in some cases of weaver syndrome but are rare in other overgrowth phenotypes. *Am. J. Hum. Genet.*, **72**, 132–143.
21. Tatton-Brown,K. and Rahman,N. (2013) The NSD1 and EZH2 overgrowth genes, similarities and differences. *Am. J. Med. Genet. C. Semin. Med. Genet.*, **163**, 86–91.
22. Wang,G.G., Cai,L., Pasillas,M.P. and Kamps,M.P. (2007) NUP98-NSD1 links H3K36 methylation to hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.*, **9**, 804–812.
23. Sanchez,R. and Zhou,M.M. (2011) The PHD finger: A versatile epigenome reader. *Trends Biochem. Sci.*, **36**, 364–372.
24. Yap,K.L. and Zhou,M.M. (2010) Keeping it in the family: Diverse histone recognition by conserved structural folds. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 488–505.
25. Pasillas,M.P., Shah,M. and Kamps,M.P. (2011) NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. *Hum. Mutat.*, **32**, 292–298.
26. He,C., Li,F., Zhang,J., Wu,J. and Shi,Y. (2013) The methyltransferase NSD3 has chromatin-binding motifs, PHD5-C5HCH, that are distinct from other NSD (nuclear receptor SET domain) family members in their histone H3 recognition. *J. Biol. Chem.*, **288**, 4692–4703.
27. Fiedler,M., Sanchez-Barrena,M.J., Nekrasov,M., Mieszczanek,J., Rybin,V., Muller,J., Evans,P. and Bienz,M. (2008) Decoding of methylated histone H3 tail by the pygo-BCL9 wnt signaling complex. *Mol. Cell*, **30**, 507–518.
28. Wang,Z., Song,J., Milne,T.A., Wang,G.G., Li,H., Allis,C.D. and Patel,D.J. (2010) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell*, **141**, 1183–1194.
29. Losson,R. and Nielsen,A.L. (2010) The NIZP1 KRAB and C2HR domains cross-talk for transcriptional regulation. *Biochim. Biophys. Acta*, **1799**, 463–468.
30. Bottomley,M.J., Stier,G., Pennacchini,D., Legube,G., Simon,B., Akhtar,A., Sattler,M. and Musco,G. (2005) NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease. *J. Biol. Chem.*, **280**, 11505–11512.
31. Gaetani,M., Matafora,V., Saare,M., Spiliotopoulos,D., Mollica,L., Quilici,G., Chignola,F., Mannella,V., Zucchelli,C., Peterson,P. *et al.* (2012) AIRE-PHD fingers are structural hubs to maintain the integrity of chromatin-associated interactome. *Nucleic Acids Res.*, **40**, 11756–11768.
32. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
33. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*, **59**, 687–696.
34. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectroscopy* **34**, 94–158.
35. Pelton,J.G., Torchia,D.A., Meadow,N.D. and Roseman,S. (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci.*, **2**, 543–558.
36. Shen,Y. and Bax,A. (2015) Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Methods Mol. Biol.*, **1260**, 17–32.
37. Sass,H.J., Musco,G., Stahl,S.J., Wingfield,P.T. and Grzesiek,S. (2001) An easy way to include weak alignment constraints into NMR structure calculations. *J. Biomol. NMR*, **21**, 275–280.
38. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
39. Yip,G.N. and Zuiderweg,E.R. (2005) Improvement of duty-cycle heating compensation in NMR spin relaxation experiments. *J. Magn. Reson.*, **176**, 171–178.
40. Johnson,B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.*, **278**, 313–352.
41. Barbato,G., Ikura,M., Kay,L.E., Pastor,R.W. and Bax,A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: The central helix is flexible. *Biochemistry*, **31**, 5269–5278.
42. Rieping,W., Habeck,M., Bardiaux,B., Bernard,A., Malliavin,T.E. and Nilges,M. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382.
43. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
44. Doreleijers,J.F., Sousa da Silva,A.W., Krieger,E., Nabuurs,S.B., Spronk,C.A., Stevens,T.J., Vranken,W.F., Vriend,G. and Vuister,G.W. (2012) CING: An integrated residue-based structure validation program suite. *J. Biomol. NMR*, **54**, 267–283.
45. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.

46. Grzesiek,S., Stahl,S.J., Wingfield,P.T. and Bax,A. (1996) The CD4 determinant for downregulation by HIV-1 nef directly binds to nef. mapping of the nef binding surface by NMR. *Biochemistry*, **35**, 10256–10261.
47. Moreira,N.H., Dolgonos,G., Aradi,B., da Rosa,A.L. and Frauenheim,T. (2009) Toward an accurate density-functional tight-binding description of zinc-containing compounds. *J. Chem. Theory Comput.*, **5**, 605–614.
48. Case,D.A., Berryman,J.T., Betz,R.M., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W. *et al.* (2015) Amber 2015. *University of California, San Francisco*.
49. Kusalik,P.G. and Svishchev,I.M. (1994) The spatial structure in liquid water. *Science*, **265**, 1219–1221.
50. Bulo,R.E., Ensing,B., Sikkema,J. and Visscher,L. (2009) Toward a practical method for adaptive QM/MM simulations. *J. Chem. Theory Comput*, **5**, 2212–2221.
51. de Vries,S.J., van Dijk,A.D., Krzeminski,M., van Dijk,M., Thureau,A., Hsu,V., Wassenaar,T. and Bonvin,A.M. (2007) HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
52. Kaminski,GA., Friesner,RA., Tirado-Rives,J. and Jorgensen,W.L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 674.
53. Peters,M.B., Yang,Y., Wang,B., Fusti-Molnar,L., Weaver,M.N. and Merz,K.M. Jr (2010) Structural survey of zinc containing proteins and the development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.*, **6**, 2935–2947.
54. Fernandez-Recio,J., Totrov,M. and Abagyan,R. (2004) Identification of protein-protein interaction sites from docking energy landscapes. *J. Mol. Biol.*, **335**, 843–865.
55. Daura,X., Antes,I., van Gunsteren,W.F., Thiel,W. and Mark,A.E. (1999) The effect of motional averaging on the calculation of NMR-derived structural properties. *Proteins*, **36**, 542–555.
56. Kwan,A.H., Gell,D.A., Verger,A., Crossley,M., Matthews,J.M. and Mackay,J.P. (2003) Engineering a protein scaffold from a PHD finger. *Structure (Camb)*, **11**, 803–813.
57. Qiu,Y., Liu,L., Zhao,C., Han,C., Li,F., Zhang,J., Wang,Y., Li,G., Mei,Y., Wu,M. *et al.* (2012) Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev.*, **26**, 1376–1391.
58. Dreveny,I., Deeves,S.E., Fulton,J., Yue,B., Messmer,M., Bhattacharya,A., Collins,H.M. and Heery,D.M. (2014) The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic Acids Res.*, **42**, 822–835.
59. Zeng,L., Zhang,Q., Li,S., Plotnikov,A.N., Walsh,M.J. and Zhou,M.M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, **466**, 258–262.
60. Cordier,F., Grubisha,O., Traincard,F., Veron,M., Delepierre,M. and Agou,F. (2009) The zinc finger of NEMO is a functional ubiquitin-binding domain. *J. Biol. Chem.*, **284**, 2902–2907.
61. Kurotaki,N., Imaizumi,K., Harada,N., Masuno,M., Kondoh,T., Nagai,T., Ohashi,H., Naritomi,K., Tsukahara,M., Makita,Y. *et al.* (2002) Haploinsufficiency of NSD1 causes sotos syndrome. *Nat. Genet.*, **30**, 365–366.
62. Musselman,C.A., Lalonde,M.E., Cote,J. and Kutateladze,T.G. (2012) Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.*, **19**, 1218–1227.
63. Ali,M., Hom,R.A., Blakeslee,W., Ikenouye,L. and Kutateladze,T.G. (2014) Diverse functions of PHD fingers of the MLL/KMT2 subfamily. *Biochim. Biophys. Acta*, **1843**, 366–371.
64. Zucchelli,C., Tamburri,S., Quilici,G., Palagano,E., Berardi,A., Saare,M., Peterson,P., Bachi,A. and Musco,G. (2014) Structure of human Sp140 PHD finger: An atypical fold interacting with Pin1. *FEBS J.*, **281**, 216–231.
65. Simpson,R.J., Cram,E.D., Czolij,R., Matthews,J.M., Crossley,M. and Mackay,J.P. (2003) CCHX zinc finger derivatives retain the ability to bind zn(II) and mediate protein-DNA interactions. *J. Biol. Chem.*, **278**, 28011–28018.
66. Wissmueller,S., Font,J., Liew,C.W., Cram,E., Schroeder,T., Turner,J., Crossley,M., Mackay,J.P. and Matthews,J.M. (2011) Protein-protein interactions: Analysis of a false positive GST pulldown result. *Proteins*, **79**, 2365–2371.
67. Malecek,K. and Ruthenburg,A. (2012) Validation of histone-binding partners by peptide pull-downs and isothermal titration calorimetry. *Methods Enzymol.*, **512**, 187–220.
68. Mackay,J.P., Sunde,M., Lowry,J.A., Crossley,M. and Matthews,J.M. (2007) Protein interactions: Is seeing believing? *Trends Biochem. Sci.*, **32**, 530–531.
69. Patel,D.J. and Wang,Z. (2013) Readout of epigenetic modifications. *Annu. Rev. Biochem.*, **82**, 81–118.
70. Spiliotopoulos,D., Spitaleri,A. and Musco,G. (2012) Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS One*, **7**, e46902.
71. Paul,F.E., Hosp,F. and Selbach,M. (2011) Analyzing protein-protein interactions by quantitative mass spectrometry. *Methods*, **54**, 387–395.
72. Gatchalian,J. and Kutateladze,T.G. (2015) PHD fingers as histone readers. In: Zhou,ZZ (ed). *Histone Recognition*. Springer, NY, pp. 27–47.
73. Zhang,H., Lu,X., Beasley,J., Mulvihill,J.J., Liu,R., Li,S. and Lee,J.Y. (2011) Reversed clinical phenotype due to a microduplication of sotos syndrome region detected by array CGH: Microcephaly, developmental delay and delayed bone age. *Am. J. Med. Genet. A.*, **155**, 1374–1378.
74. Jennings,B.H., Pickles,L.M., Wainwright,S.M., Roe,S.M., Pearl,L.H. and Ish-Horowicz,D. (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol. Cell*, **22**, 645–655.