J Biomol NMR (2013) 56:183–190
DOI 10.1007/s10858-013-9726-x

NMR STRUCTURE NOTE

NMR structure note: PHD domain from death inducer obliterator protein and its interaction with H3K4me3

Clara M. Santiveri · M. Flor García-Mayoral · José M. Pérez-Cañadillas · M. Ángeles Jiménez

Received: 6 March 2013 / Accepted: 2 April 2013 / Published online: 12 April 2013

Biological context

The plant homeodomain (PHD) modules are small 50–80 amino acid zinc fingers present in many nuclear proteins, which recognise histone post-translational modifications, i.e. lysine methylation and acetylation (Li and Li 2012; Musselman et al. 2012; Sanchez and Zhou 2011; Baker et al. 2008). These modifications play an essential role in the regulation of transcription, activation or repression depending on the nature and extent of the modification and on the target lysine. Misreading of these epigenetic marks has been related with many human pathological states, such as cancer, immunological and neurological diseases (Musselman et al. 2012; Baker et al. 2008).

The death inducer obliterator (*Dido*) gene encodes three protein isoforms of different lengths. The longest and most broadly expressed, Dido3, is a nuclear protein that associates to the spindle pole in mitosis and to the synaptonemal complex in meiosis. Alterations in the expression of the *Dido* gene have been related to myeloid neoplasms in humans (Fütterer et al. 2005). Based on pull-down assays, the N-terminal region of murine Dido3 has been reported to associate to histone H3 (Prieto et al. 2009). Histone recognition requires the PHD motif present in all Dido isoforms at their common N-terminal region (Fig. 1a; Prieto et al. 2009). It is noticeable that the PHD domain sequence in *Dido* genes from different organisms is completely conserved, whilst the overall identities lie in the range 60–96 %. Surface plasmon resonance experiments indicated that, in vitro, the Dido PHD domain is able to bind histone H3-derived peptides. The affinity is higher for the peptide with trimethylated-lysine 4 (H3K4me3) than for its non-methylated counterpart. Dido PHD domain was shown to recognise H3K4me3 also in vivo, and the methylation state of lysine 4 seems to be involved in the cellular localization of Dido3 (Prieto et al. 2009). Thus, knowledge of the molecular basis for the interaction between the PHD domain of Dido and histone H3K4me3 would improve current understanding of the biological roles played by Dido.

With this aim in mind we proceeded to determine the structure of the PHD domain of Dido (DidoPHD), residues 265–322 in humans (Fig. 1a), and to map its interaction with a 12-residue H3K4me3 histone peptide (Fig. 1b).

Methods and results

Protein expression and purification

A synthetic gene with codon usage optimized for *E. coli* corresponding to the PHD domain (265–322) of human Dido was sub-cloned into a modified pET28 vector (Novagen) containing an N-terminal GST fusion and a TEV protease site. This plasmid was transformed into *E. coli* BL21 (DE3) cells (Novagen) which were grown in

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9726-x) contains supplementary material, which is available to authorized users.

C. M. Santiveri · M. F. García-Mayoral · J. M. Pérez-Cañadillas · M. Á. Jiménez (✉)
Instituto de Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain
e-mail: majimenez@iqfr.csic.es
URL: http://rmnpro.iqfr.csic.es/

Present Address:
C. M. Santiveri
Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

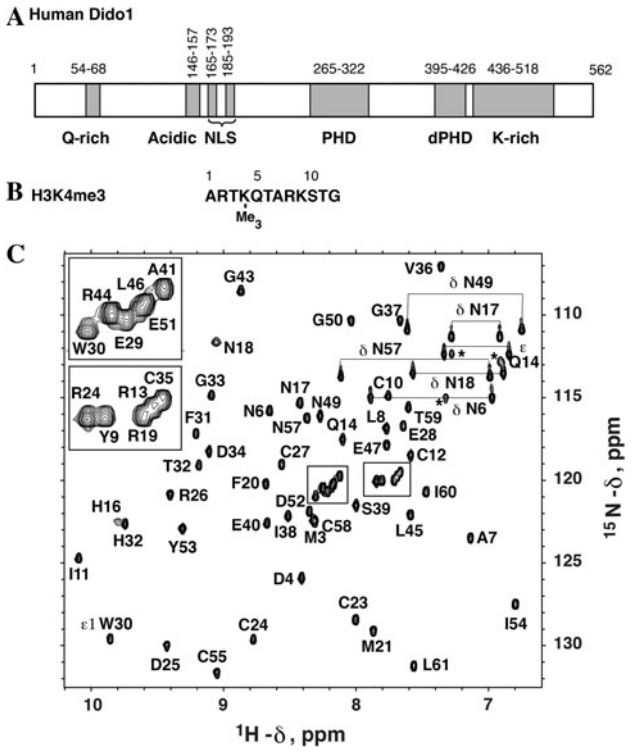

Fig. 1 **a** Scheme of the N-terminal region common to the three isoforms of human Dido showing the domains predicted from sequence analysis: Q-rich region, acidic region, a bipartite nuclear localization signal motif (NLS), PHD motif, a short Zn-finger motif denoted as dPHD, and a K-rich region (García-Domingo et al. 1999; Rojas et al. 2005). These motifs are coloured in *grey*, labelled at the *bottom*, and their boundaries are indicated on *top*. **b** Sequence of the histone peptide H3K4me3 used in this work. **c** 2D [^{1}H,^{15}N]-HSQC spectrum of [^{15}N,^{13}C]-DidoPHD 0.4 mM in 10 mM potassium phosphate pH 6.5, 100 μM TCEP, 10 μM DSS, 0.01 % sodium azide (H_2O/D_2O 9:1 v/v). The *boxed crowded* regions are shown expanded as *insets*. Residues are numbered including the cloning tag. The cross-peak corresponding to the W indolic NH is labelled (ε1), the two cross-peaks observed for the side chain amide group of N and Q residues are linked by a line and labelled (δ for N and ε for Q), and the folded NεH cross-peaks of R are indicated by *asterisks*

LB medium supplemented with 30 μg/l of kanamycin (Sigma-Aldrich) to get non-labelled DidoPHD and in K-MOPS-derived minimal medium (Neidhardt et al. 1974) supplemented with $^{15}NH_4Cl$ (1 g/l) and ^{13}C-glucose (4 g/l) to get [^{13}C,^{15}N]-DidoPHD. In both cases, the cultures were grown at 37 °C up to $OD_{600} = 0.3$, supplemented with $ZnCl_2$ (10 μM final concentration), cooled to 20 °C, further grown to $OD_{600} = 0.6$ and induced with 0.5 mM IPTG (Sigma-Aldrich) overnight. Cultures were harvested by centrifugation at 3,000*g* for 30 min and pellets were re-suspended in PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 pH 7.4) containing 5 mM DTT and 1 tablet of EDTA-free protease inhibitor cocktail (Roche) per 50 ml and lysed by French-press. The bacterial lysates were centrifuged at 40,000*g* for 30 min and soluble fractions were incubated with 1/5 of volume of Glutathione Sepharose 4 Fast Flow (GE Healthcare) on ice (rocking) for 1 h. The mixture was applied to a gravity column and the resin washed with 5 column volumes of binding buffer. The fusion protein was eluted with 50 mM Tris–HCl, 10 mM reduced glutathione pH 8.0 buffer, dialyzed against 20 mM Tris–HCl pH 8.0 buffer and simultaneously cleaved with home-made TEV protease (~1 mg/ml) overnight at 4 °C. The sample was loaded into two 5 ml SP and mono Q columns (GE Healthcare) arranged in series. TEV protease bound to the SP column while GST tag and DidoPHD bound to the mono Q column. The proteins were eluted with a linear salt gradient (from 0 to 1 M NaCl) in a buffer containing 20 mM Tris–HCl pH 8.0 and 1 mM DTT and further purified by gel filtration chromatography with a Superdex 75 column (GE Healthcare) in 25 mM NaCl, 100 μM DTT, 10 mM potassium phosphate buffer pH 6.5. Fractions containing pure DidoPHD protein, as checked by SDS-PAGE, were pooled, concentrated and dialyzed against NMR buffer (10 mM potassium phosphate pH 6.5, 100 μM TCEP) to yield 0.4–0.6 mM protein concentration. The pH value was checked with a glass microelectrode and was not corrected for isotope effects.

NMR spectra acquisition

NMR spectra were acquired on Bruker AV 600 and 800 MHz spectrometers both equipped with z-gradient cryoprobes. A methanol sample was employed to calibrate the NMR probe temperature. ^{1}H chemical shifts were referenced to the internal sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS), and ^{13}C and ^{15}N chemical shifts indirectly referenced by multiplying the ^{1}H spectrometer frequency assigned to 0 ppm by 0.251449530 and 0.101329118, respectively (Markley et al. 1998). All NMR spectra were processed using the TOPSPIN software (Bruker Biospin, Karlsruhe, Germany) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, Sparky 3, University of California, San Francisco, USA). 2D [^{1}H–^{15}N]-HSQC, 2D [^{1}H–^{13}C]-HSQC and 3D HNCO, HNCA, HN(CA)CO, CBCA(CO)NH, CBCANH, HBHA(CO)NH, HNHA, and HCCH–TOCSY experiments were acquired using a [^{13}C, ^{15}N]-DidoPHD sample, and 2D homonuclear [^{1}H,^{1}H]-COSY, 60 ms [^{1}H,^{1}H]-TOCSY and 80 ms [^{1}H,^{1}H]-NOESY experiments were recorded using a non-labelled DidoPHD sample.

NMR chemical shift assignment

The [^{1}H–^{15}N]-HSQC spectrum acquired for [^{13}C,^{15}N]-DidoPHD in aqueous solution at 25 °C (Fig. 1c) showed cross-peaks for all the amide groups, except for the N-terminal GS sequence that belongs to the cloning tag. A

near-complete assignment of all backbone (^{15}N, ^{1}HN, $^{1}H_{\alpha}$, $^{13}C_{\alpha}$ and ^{13}C') and $^{13}C_{\beta}$ atoms was achieved by analysing series of 3D NMR spectra (CBCANH, CBCA(CO)NH, HBHA(CO)NH, HNCO, HN(CA)CO, HNCA and HNHA). These analyses also allowed us to identify the resonances for the W imino group, the six N/Q amide side chain groups, including the corresponding $^{13}C\gamma$ and $^{13}C\varepsilon$ carbonyl groups, and the $^{1}H\varepsilon$ and $^{15}N\varepsilon$ atoms of five R (13, 19, 26, 44, 48) out of six present in the DidoPHD sequence. From the analysis of 3D HCCH–TOCSY, 2D [^{1}H–^{1}H]-TOCSY, 2D [^{1}H–^{1}H]-COSY and 2D [^{1}H–^{1}H]-NOESY spectra the assignment was propagated to other ^{1}H and ^{13}C side chain resonances, including aromatic rings and P residues. The three X–P bonds present in DidoPHD are in the *trans* conformation as indicated by the chemical shift difference between $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ carbons, $\Delta\delta_{C\beta-C\gamma} =$ 4.3–5.6 ppm (Schubert et al. 2002), and by the sequential NOEs between the H_{δ} protons of P5, P15 and P56, and the H_{α} of their preceding residue observed in the 2D [^{1}H–^{1}H]-NOESY spectrum.

Concerning the eight Cys residues, their $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts indicate that all of them are reduced, and that C23 very probably has a free sulphydryl group, whilst the other seven Cys residues coordinate a Zn^{2+} ion (Kornhaber et al. 2006). All these chemical shifts were deposited at the BioMagResBank (http://www.bmrb.wisc.edu/; accession code BMRB-18963).

NMR structure calculation and Zn^{2+} binding sites

Structure calculations were performed using CYANA 2.1 and running the standard iterative protocol for automatic NOE assignment consisting of seven cycles of combined automated assignment of NOE cross-peaks and structure calculation of 100 conformers per cycle (Güntert 2004). The upper limit distance constraints were derived from the integrated cross-peaks of an 80-ms 2D [^{1}H–^{1}H]-NOESY spectrum acquired for non-labelled DidoPHD, and the list of all the ^{1}H chemical shifts plus the ^{13}C chemical shifts for P residues, which are necessary to determine the X–P conformation. The NOE cross-peaks were integrated using Sparky. The dihedral ϕ and Ψ angle restraints were obtained from $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}C' and ^{15}N chemical shifts using the program TALOS (Cornilescu et al. 1999).

In the initial structural ensemble obtained by the above protocol, the side chains of two four-residue groups, i.e. C24, C27, C55 and C58 in a putative CCCC motif, and C10, C12, H32, and C35 in a CCHC motif, are positioned adequately to coordinate Zn^{2+} ions. In the subsequent structure calculations the two Zn^{2+} ions and the upper and lower distance limits required for canonical Zn^{2+} coordination were added (Table ST3 at Supp. Info). To that end, a modified Cys residue with a bound Zn^{2+} (CysZ) was included in the CYANA residue library. Since a single CysZ residue is required for each Zn^{2+}-binding motif, we checked that the final structure ensembles did not depend on the CysZ position within the CCHC (10/12/35) and CCCC (24/27/55/58) motifs. In the case of the CCHC motif, we also examined whether the metal ion is coordinated to H32 by the $N_{\varepsilon 2}$ nitrogen in an Hδ tautomer or by the $N_{\delta 1}$ nitrogen in an Hε tautomer. The fact that the $N_{\delta 1}$ nitrogen of H32 is closer to the sulfur atoms of the Cys residues of the CCHC motif than the $N_{\varepsilon 2}$ nitrogen in the initial DidoPHD structure suggested that H32 coordinates Zn^{2+} by the $N_{\delta 1}$ nitrogen. This was confirmed by the structure calculations forcing Zn^{2+} coordination to the $N_{\varepsilon 2}$ nitrogen of H32, which led to more violations than those forcing Zn^{2+} coordination to the $N_{\delta 1}$ nitrogen. Also, the ^{13}C chemical shifts of $C_{\delta 2}$ and $C_{\varepsilon 1}$ of H32 (118.3 and 138.6 ppm, respectively) match closely those corresponding to $N_{\delta 1}$-coordination of the Zn^{2+} ion (Barraud et al. 2012).

The final structure calculation was done following the standard automatic protocol of CYANA 2.1, as described above, plus the upper and lower distance limit restraints required for Zn^{2+} coordination. In the seventh cycle of this protocol, stereospecific assignments are automatically done on the basis of the NOEs involving them and the corresponding distances in the structure. In this way, the $H_{\alpha 2}/H_{\alpha 3}$ protons of G33 and G43 residues, the $H_{\beta 2}/H_{\beta 3}$ protons of 22 residues (D4, L8, Y9, C10, C12, Q14, H16, F20, M21, C23, C24, D25, R26, E29, F31, H32, C35, R44, C55, P56, N57, C58), the $H_{\gamma 2}/H_{\gamma 3}$ protons of 5 residues (I11, P15, M21, I22, I54), the $H_{\delta 2}/H_{\delta 3}$ protons of the three P residues (P5, P15 and P56), the γ1 and γ2 methyl groups of V36, the δ1 and δ2 methyl groups of L8, and the amide side chain protons of N49 and N57 were stereo-specifically assigned. The ensemble of the 20 DidoPHD conformers with the lowest target functions resulting from this final calculation was deposited at the PDB with accession code 2m3h (Fig. 2; Table 1).

Titration of DidoPHD domain with a histone H3 peptide trimethylated at Lys 4 (H3K4me3)

To understand how DidoPHD domain recognises trimethylated-lysine in histone H3 peptides, a [^{13}C, ^{15}N]-DidoPHD sample was titrated with increasing amounts of peptide H3K4me3 (Fig. 1b; synthesised by Peptide Protein Research Ltd). 2D [^{1}H–^{15}N]-HSQC and [^{1}H–^{13}C]-HSQC spectra were acquired at 25 °C at protein/peptide ratios of 1:0, 1:0.8, 1:1.6, 1:2.4, 1:3.2 and 1:4. Many cross-peaks in the [^{1}H–^{15}N]-HSQC (Fig. 3a) and [^{1}H–^{13}C]-HSQC spectra are shifted upon titration with peptide H3K4me3. Therefore, assignment of the bound DidoPHD was confirmed by analyses of 2D CBCANH, CBCA(CO)NH, HNCO,

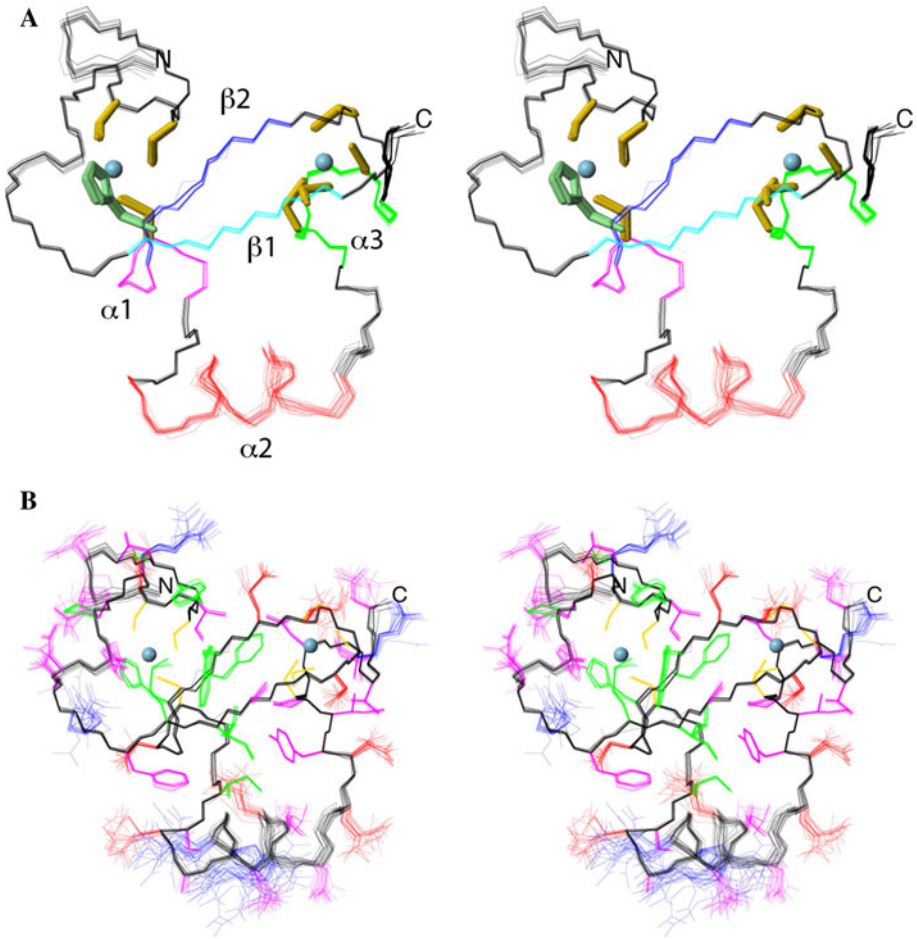

Fig. 2 NMR solution structure ensemble of DidoPHD. Stereo views of the backbone atoms overlay for residues 4–61. In **a**, the strands β1 and β2 are coloured in *cyan* and *blue*, respectively, and the helices α1, α2, and α3 in *magenta*, *red*, and *green*. The Zn^{2+}-binding sites are highlighted by displaying the Zn^{2+} ions as *light blue spheres*, and the Zn^{2+}-bound side chains in *neon* with Cys and His residues coloured in *yellow* and *light green*, respectively. In **b**, all the backbone atoms are displayed in *black*, and the side chains in *blue* if positively charged (Lys, Arg), *red* if negatively charged (Asp, Glu), *yellow* for the Cys residues, *green* for the core residues (those with buried side chains; total ASA ≤20 %), and *magenta* for all others. In both panels, the N- and C-termini are labelled with "N" and "C", respectively

HN(CA)CO and HNCA spectra recorded at the last titration point. In this way, we achieved an almost complete assignment of ^{1}H, ^{13}C and ^{15}N resonances of the DidoPHD protein in its complex with peptide H3K4me3, which were deposited at BioMagResBank (http://www.bmrb.wisc.edu/; accession code BMRB-19074). A histogram showing the average of the changes experienced by the ^{1}H and ^{15}N of the amide groups as a function of sequence is shown in Fig. 3b. The residues showing the highest perturbations suggest that H3K4me3 binds to DidoPHD using the canonical PHD binding site.

Model building of DidoPHD/H3K4me3 complex

To better visualise how this interaction occurs, we proceeded to build a model of the complex DidoPHD/H3K4me3 by using the Haddock-webserver (http://haddock.chem.uu.nl/services/HADDOCK/haddockserver-expert.html). This program requires input coordinates for both, the protein and the peptide, and a definition of their interacting residues, classified as active and passive. For the protein, we used the PDB coordinates determined in this work for the solution structure of the free DidoPHD domain structure, and the active and passive residues were derived from the titration of the labeled protein with the unlabeled H3K4me3 peptide. Thus, the residues with average HN chemical shift perturbations, $\Delta\delta^{av}$, larger than 0.25 ppm (D4, Y9, N17, R19, M21, C23, R26, W30, G33, S39, D52, and Y53; Fig. 3b) were considered as active. Considering that the most strongly perturbed residues are localized at segment 16–34, the residues at this region with $\Delta\delta^{av}$ values less than 0.25 ppm were included as passive (H16, N18, F20, I22, C24, D25, C27, E28, E29, F31, H32 and D34). Most of them have $\Delta\delta^{av} > 0.20$ ppm, or present large perturbations at the ^{1}H and ^{13}C chemical shifts of their side chains, as in F20 and I22. The residue N6 with $\Delta\delta^{av} > 0.20$ ppm was also included as passive. These residues possess ASA values >20 %, except for M21, I22, G33, C24, C27 and the aromatic residues. In a 2D TOCSY spectrum acquired for a [^{13}C, ^{15}N]-DidoPHD/non-labelled H3K4me3 1:2 sample of the complex, signals corresponding to peptide residues 8–12 remained unchanged relative to the free peptide; hence we built the H3K4me3 peptide with the first 8 residues of the sequence and defined all these residues as active. The cross-peaks for the perturbed residues, in particular those of the trimethylated K4, could not be identified, which precluded the finding of

Table 1 Structural statistics for the ensemble of the 20 lowest target function structures of DidoPHD

Number of NOE distance restraints	
Intraresidue ($i - j = 0$)	270
Sequential ($\|i - j\| = 1$)	273
Medium range ($1 < \|i - j\| < 5$)	169
Long-range ($\|i - j\| \geq 5$)	378
Total number	1,090
Averaged total number per residue	17.9
Number of distance restraints for Zn^{2+} ligand[a]	
Upper limit	23
Lower limit	23
Number of dihedral angle constraints	
Number of restricted ϕ angles	54
Number of restricted ψ angles	48
Total number	102
Average maximum violations per structure	
Distance (Å)	0.08 ± 0.01
Dihedral angle (°)	1.5 ± 0.1
Deviations from ideal geometry	
Bond length (Å)	0.001
Bond angle (°)	0.2
Pairwise rmsd (Å)	
All residues (4–61)[b]	
Backbone atoms	0.3 ± 0.1
All heavy atoms	1.0 ± 0.1
Ramachandran plot (%)	
Residues in most favoured regions	78.8
Residues in additional allowed regions	21.2
Residues in generously allowed regions	0
Residues in disallowed regions	0

[a] Table ST3 at Supp. Info lists the distance restraints included to improve the geometry of Zn^{2+} coordination

[b] The cloning tag is excluded

intermolecular NOEs. It is feasible that the trimethylammonium signal was not detectable due to motions within the aromatic pocket in the slow to intermediate NMR time scale leading to line broadening. Since the isolated peptide is mainly random coil in aqueous solution (see Supp. Info), we took the peptide coordinates from the first deposited model of the solution structure of the complex of an H3K4me3 peptide with TAF3PHD (pdb code: 2k17; van Ingen et al. 2008). Peptide and protein termini were charged with Pymol (DeLano, W. L., Pymol Molecular Graphics System, 2006, DeLano Scientific, San Carlos, CA, USA). With these data and fixing H16 and the Zn^{2+}-coordinated H32 to the Hε tautomer, as in free DidoPHD (see above), 200 complex models were generated using the Haddock Expert Interface. To improve the convergence of the peptide backbone conformation in the complex structures, characteristic hydrogen-bond restraints for antiparallel β-sheet were introduced connecting residues R2-K4me_3-T6 of the peptide with residues C23-M21-R19 of the protein, as commonly observed in PHD/histoneH3-peptide complexes (Musselman et al. 2012; Sanchez and Zhou 2011). The majority of the resulting model complex structures were grouped in a unique cluster with the best Haddock score of −96 kcal/mol (Z score, −1.0) and restraint violation energy of 28.2 kcal/mol. The average surface buried at the interface is 1,291 $Å^2$. The backbone rmsd for the octapeptide (A1-R8) in the family of the 20 best structures of the complex is 1.16 Å. When only the core interacting residues (A1-T6) are considered this value reduces to 0.96 Å (see also Fig. SF1 at Supp. Info).

Discussion and conclusions

The solution structure of DidoPHD

The structural ensemble of DidoPHD (Fig. 2) was examined using MOLMOL (Koradi et al. 1996) and PROMOTIF (Hutchinson and Thornton 1996). It is well-defined and all the residues are either in the most favoured or allowed regions of the Ramachandran map (Table 1). Four residues exhibit positive ϕ dihedral angles: R13, E28, G37 and G50. Excluding G, A and P (9 in total), the only residues whose side chains are not ordered (a χ_1 angular rmsd $\leq 30°$ is considered indicative of an ordered side chain) are S2 and M3, both belonging to the flexible N-terminus, and N6, R19, E28, E40, R42, R44 and D52, all of them solvent-exposed (ASA $\geq$ 20 %).

DidoPHD structure (Fig. 2) shows a 4:4 β-hairpin, formed by two antiparallel β-strands (β1, 21–24, and β2, 29–32) connected by a type I β-turn, and three α-helices (α1, 33–36; α2, 40–49, and α3, 54–59), two of them very short. Also, it displays the characteristic interleaved topology of the two Zn^{2+} coordinating motifs (CCHC and CCCC; Fig. 2a). A cleft between the strand β1 and the longest helix α2 can also be distinguished (Fig. 2a). This structure presents a Dali (Holm and Sander 1993) Z score of 8.1 and an rmsd of 1.2 Å with the NMR structure of a 76-residue murine PHD construction (pdb code: 1wem). Fig. 3c shows the sequence alignment of DidoPHD with PHD domains representative of different protein families, selected by high Dali Z score >4.4, rmsd < 2.5 Å, and high sequence similarity. The region with the highest variability corresponds to the long helix α2 in DidoPHD, which, indeed, is a loop or a very short helix in many PHD domains. A peculiarity of the DidoPHD sequence is an additional cysteine preceding the first Cys of the CCCC Zn^{2+}-binding site, only shared by two other PHD fingers, those of the proteins ALFL4 and MLL5 (Fig. 3c). The functional relevance, if any, of this Cys residue that is reduced and has its side chain quite solvent-exposed (side chain ASA is 56 %) remains to be established.

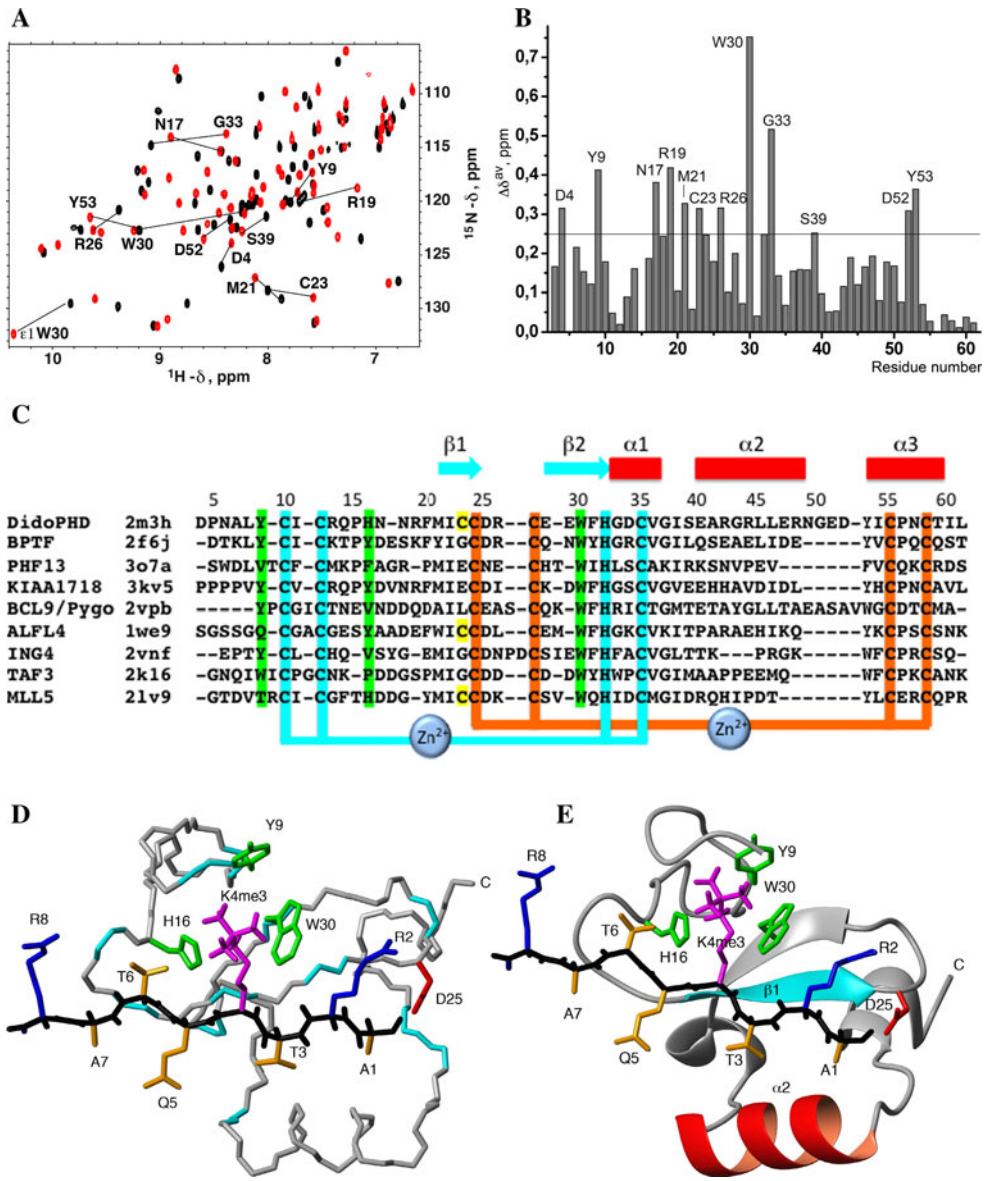

Fig. 3 Interaction of DidoPHD with H3K4me3. **a** Overlay of the 2D [^{1}H–^{15}N]-HSQC spectra acquired for the free [^{13}C,^{15}N]-DidoPHD (in *black*) and at a [^{13}C,^{15}N]-DidoPHD/unlabelled H3K4me3 1:4 ratio (in *red*). Cross-peaks for the residues with $\Delta\delta^{av} > 0.25$ ppm are connected by a line and labelled. **b** Histogram of averaged amide ^{1}H and ^{15}N chemical shift perturbations ($\Delta\delta^{av} = \{[(\Delta\delta_{1H})^2 + (\Delta\delta_{15N}/5)^2]/2\}^{1/2}$, ppm). A *horizontal line* indicates $\Delta\delta^{av} = 0.25$ ppm, and the residues above this limit are labelled. **c** Sequential alignment of DidoPHD with representative PHD domains (pdb codes are indicated). Among those with high sequence similarity to DidoPHD, we selected those with high Dali score and small rmsd versus DidoPHD value (Dali Z score and rmsd for the aligned residues are, respectively, 8.1 and 1.6 Å for 2f6j, 7.4 and 1.7 Å for 3o7a, 7.3 and 1.6 Å for 3kv5, 7.0 and 1.7 Å for 2vpb, 6.3 and 1.8 Å for 1we9, 5.5 and 1.8 Å for 2vnf, 4.9 and 2.2 Å for 2k16, and 4.4 and 2.4 Å for 2lv9). Secondary structure is indicated on *top*, the β-strands as *cyan arrows* and the helices as *red rectangles*. The CCHC and CCCC Zn^{2+}-binding sites are coloured in *cyan* and *orange*, respectively. The aromatic cage of the canonical H3K4me3 binding site is coloured in *green*. The rare reduced cysteine is shown in *yellow*. **d**, **e** Model structure of the DidoPHD/H3K4me3 complex with the protein backbone displayed in *grey neon* in **d** and as a *ribbon* in **e**. Strand β1 is coloured in *cyan*, and helix α2 in *red*. The peptide backbone is shown in *black*. The DidoPHD side chains interacting with peptide H3K4me3 are highlighted and labelled. The aromatic residues Y9, H16 and W30 are coloured in *green*, and D25 in *red*. Side chains for all residues in peptide H3K4me3 are displayed. The side chain of the trimethylated K4 is shown in *magenta*, R2 and R8 in *blue*, and all the other in *orange*. The protein residues with $\Delta\delta^{av} > 0.25$ ppm are coloured in *cyan* onto the protein backbone in **d**

Interactions of DidoPHD with H3K4me3

The model of the complex structure is mainly maintained by electrostatic and hydrophobic interactions. In more than one-third of the structures out of 20 selected for analysis, the side chain of A1 locates at distances shorter than 5 Å from the I22, C23, L46, and Y53 side chains, with the charged N-terminus contacting the carboxylate of D25. The

side chain of R2 is solvent-exposed (ASA = 28 %) and not well-ordered in the model, however, in almost half of the conformers of the complex family it is located in the proximity of the carboxylate of D25, similarly to many other solved complexes, while in a few others it is closer to E47. The side chain of the trimethylated K4 appears clustered into two main different orientations. The most populated orientation places this side chain in the proximity of W30, a conserved residue involved in π-cation interactions. The other residue that is typically involved in these interactions is a conserved tyrosine that forms with the tryptophan an aromatic cage that stabilizes the charged trimethyl-amino group in other PHD complexes (Li and Li 2012; Sanchez and Zhou 2011). Interestingly, our model reveals distances to the conserved tyrosine, Y9, slightly larger, however, a non-conserved histidine residue, H16, is located nearby opposite to the W30 delimiting the aromatic cage on both sides of the $K4me_3$ side chain with Y9 at the base (Fig. 3d, e).

Strikingly, a few model complexes show a completely different orientation of the side chain of K4me3 pointing mid-distance between the negative charge of E47 and the aromatic ring of F20, two non-conserved residues with which electrostatic and π-cation interactions can also take place (see Fig. SF1 at Supp. Info). E47 lies at the middle of helix α2, a region with high sequence and structural variability among PHD domains. Interestingly, the ^{1}H and ^{13}C chemical shift perturbations for the aromatic ring of F20 are relatively large (absolute values for $\Delta\delta^{complex-free}$ are in the range 0.07–0.22 ppm for the ^{1}H, and about 1 ppm for the ^{13}C; see Table ST4 at Supp. Info). The chemical shifts of the E47 side chain are less affected, but one H_γ proton exhibits a $\Delta\delta^{complex-free}$ of 0.1 ppm. On the other hand, the aromatic residue F20 is not generally conserved in PHD domains, but it is present in some of those structurally similar to DidoPHD (see above; Fig. 3c).

Conclusions

The PHD domain of Dido, a nuclear protein involved in mitosis and meiosis (Fütterer et al. 2005; Prieto et al. 2009), has identical sequences in the *Dido* gene from all organisms reported up to now. This DidoPHD module adopts a globular structure by itself, which displays the topology characteristic of PHD domains, with a long helix α2 (Fig. 2). This domain interacts with a histone H3-derived peptide with the trimethylated K4 inserted into a typical aromatic cage, formed in this case by residues Y9, H16, and W30 (Fig. 3c, d). In addition, our results point out towards the potential existence of a second binding sub-site for the trimethylated K4, which consists of F20, a very poorly conserved residue, and E47, a non-conserved residue (Fig. 3c). On the whole, our data constitute a basis for a molecular understanding of the Dido functionality.

Acknowledgments We thank C. López, Dr. D. Pantoja-Uceda, P. Rico-Lastres and L. de la Vega at the Instituto de Química Física Rocasolano, CSIC, for technical assistance, and to Dr. K. van Wely at the Centro Nacional de Biotecnología, CSIC, for providing us the synthetic gene. This work was supported by grant S2010/BMD-2305 from the Spanish Comunidad de Madrid, and projects CTQ2011-22514 and CTQ2011-26665 from the Spanish MINECO.

References

Baker LA, Allis CD, Wang GG (2008) PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. Mutat Res 647:3–12. doi:10.1016/j.mrfmmm.2008.07.004

Barraud P, Schubert M, Allain FH (2012) A strong ^{13}C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins. J Biomol NMR 53:93–101. doi:10.1007/s10858-012-9625-6

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302. doi:10.1023/A:1008392405740

Fütterer A, Campanero MR, Leonardo E, Criado LM, Flores JM, Hernández JM, San Miguel JF, Martínez-A C (2005) Dido gene expression alterations are implicated in the induction of hematological myeloid neoplasms. J Clin Invest 115:2351–2362. doi:10.1172/JCI24177

García-Domingo D, Leonardo E, Grandien A, Martínez P, Albar JP, Izpisúa-Belmonte JC, Martínez-A C (1999) DIO-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development. Proc Natl Acad Sci USA 96:7992–7997. doi:10.1073/pnas.96.14.7992

Güntert P (2004) Automated NMR protein structure calculation. Prog Nucl Magn Res Spect 43:105–125. doi:10.1016/S0079-6565(03)00021-9

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233:23–38. doi:10.1006/jmbi.1993.1489

Hutchinson EG, Thornton JM (1996) PROMOTIF-a program to identify and analyze structural motifs in proteins. Protein Sci 5:212–220. doi:10.1002/pro.5560050204

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–58. doi:10.1016/0263-7855(96)00009-4

Kornhaber GJ, Snyder D, Moseley HN, Montelione GT (2006) Identification of zinc-ligated cysteine residues based on 13Calpha and 13Cbeta chemical shift data. J Biomol NMR 34:259–269. doi:10.1007/s10858-006-0027-5

Li Y, Li H (2012) Many keys to push: diversifying the 'readership' of plant homeodomain fingers. Acta Biochim Biophys Sin 44:28–39. doi:10.1093/abbs/gmr117

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids—(IUPAC Recommendations 1998). Pure Appl Chem 70:117–142. doi:10.1351/pac199870010117

Musselman CA, Lalonde ME, Côté J, Kutateladze TG (2012) Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol 19:1218–1227. doi:10.1038/nsmb.2436

Neidhardt FC, Bloch PL, Smith DF (1974) Culture medium for enterobacteria. J Bacteriol 119:736–747

Prieto I, Kouznetsova A, Fütterer A, Trachana V, Leonardo E, Alonso Guerrero A, Cano Gamero M, Pacios-Bras C, Leh H, Buckle M, Garcia-Gallo M, Kremer L, Serrano A, Roncal F, Albar JP, Barbero JL, Martínez-A C, van Wely KH (2009) Synaptonemal complex assembly and H3K4me3 demethylation determine DIDO3 localization in meiosis. Chromosoma 118:617–632. doi:10.1007/s00412-009-0223-7

Rojas AM, Sanchez-Pulido L, Fütterer A, van Kelly KHM, Martínez-A C, Valencia A (2005) Death inducer obliterator protein 1 in the context of DNA regulation. Sequence analyses of distant homologues point to a novel functional role. FEBS J 272:3505–3511. doi:10.1111/j.1742-4658.2005.04759.x

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. Trends Biochem Sci 36:364–372. doi:10.1016/j.tibs.2011.03.005

Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. J Biomol NMR 24:149–154. doi:10.1023/A:1020997118364

van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers HT, Boelens R (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure 16:1245–1256. doi:10.1016/j.str.2008.04.015