Crystal structures of the Tudor domains of human PHF20 reveal novel structural variations on the Royal Family of proteins

Melanie A. Adams-Cioaba, Zhihong Li, Wolfram Tempel, Yahong Guo, Chuanbing Bian, Yanjun Li, Robert Lam, Jinrong Min

aStructural Genomics Consortium, University of Toronto, 101 College St., Toronto, Ontario, Canada M5G 1L7
bDepartment of Physiology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

ABSTRACT

The human PHD finger protein 20 (PHF20) is a putative transcription factor. While little is known about its cognate cellular role, antibodies against PHF20 are present in sera from patients with hepatocellular carcinoma, glioblastoma and childhood medulloblastoma. PHF20 comprises two N-terminal Tudor domains, a central C2H2-link zinc finger domain and a C-terminal zinc-binding PHD domain, and is a component of some MLL methyltransferase complexes. Here, we report the crystal structures of the N-terminal Tudor domains of PHF20 and highlight the novel structural features of each domain. We also confirm previous studies suggesting that the second Tudor domain of PHF20 exhibits preference for dimethylated histone substrates.

© 2012 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

The recognition of post-translational modifications by various protein domains represents an important mechanism for cellular regulation. The ability of members of the Royal Family of proteins, including the Tudor domain subfamily, in binding methylated protein substrate has been well documented (reviewed in [1–6]). However, much remains to be elucidated with regard to precise mechanisms by which such interactions influence the processes of transcription, translation and RNA splicing.

The tails of histones H3, H4, H2A and H2B make contacts with the poly-anionic backbone of core nucleosomal DNA, thereby contributing to nucleosome stability and in turn the high order chromatin structure [7]. The regulation of chromatin structure and the resultant transcriptional activity within a given region of chromatin is influenced by the post-translational modifications of histone tails [8–10]. In particular, histones H3 and H4 are subject to numerous post-translational modifications at multiple sites, including the mono-, di- and tri-methylation of lysine residues distributed along the tails. Catalyzed by large complex-forming methyltransferases, the deposition of such methyl marks has been associated with states of active or repressed transcription and the recognition of such marks appears important in targeting large multicomponent protein complexes to their respective sites [11].

The PHD finger protein 20 (PHF20/CLEA2/HCAS8) is a component of some mixed-lineage leukemia (MLL) methyltransferase complexes with the core components MLL, ASH2L, WDR5 and RBBP5 [12]. Clinically, antibodies against PHF20 are present in sera from patients with hepatocellular carcinoma, glioblastoma and childhood medulloblastoma [13–15]. It is one of few immunogenic antigens identified via ELISA for gliomas and exhibits the most frequent seroreactive response from glioma patients [15].

PHF20 comprises two N-terminal Tudor domains, a central C2H2-link zinc finger domain and a C-terminal zinc-binding PHD domain (Fig. 1). Although little is known about its cellular role, the domain organization of PHF20 and the association of PHF20 with MLL core complexes suggest it to function as a transcription factor. Previous studies have indicated that the second Tudor domain is capable of binding methylated residues on histone tails [16], while no such function has been ascribed to the first Tudor domain in this work. We report the crystal structures of the two N-terminal Tudor domains of PHF20. We show that both Tudor domains possess novel structural features, which may underlie their respective cellular functions. We have also explored the potential interactions of the respective domains with various histone-derived peptides by fluorescence polarization assays and
revealed that the second Tudor domain of PHF20 displays preference for dimethylated histones substrates.

2. Materials and methods

2.1. Expression, purification and crystallization of PHF20 Tudor domains

Two fragments of human PHF20 (residues 1–81 and 83–150) covering the first and second Tudor domain, respectively, were subcloned into a pET-28a-MHL vector. The recombinant protein was over-expressed at 18°C in an N-terminal His₆-tagged protein in Escherichia coli BL21 (DE3) Codon plus RIL (Strategene). Following purification via Ni–NTA column and size exclusion chromatography (Superdex 75, GE Healthcare), each protein was concentrated to 20 mg/ml in 20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT.

Purified PHF20 Tudor domain 1 (residues 1–81) was crystallized in a buffer containing 25% PEG 3350, 0.2 M ammonium acetate, 0.1 M bis–tris pH 5.5 using the sitting drop vapor diffusion method at 18°C. The crystal appeared overnight. Prior to freezing, the crystals were soaked in a drop containing mother liquor plus 10–15% glycerol.

Purified PHF20 Tudor domain 2 (83–150) was crystallized in 25% PEG 8000, 0.2 M sodium chloride, 0.1 M Tris, pH 8.5 using sitting drop vapor diffusion at 18°C. The crystal appeared overnight. Prior to freezing, the crystals were soaked in a drop containing mother liquor plus 10–15% glycerol.

2.2. Fluorescence polarization binding assay

All peptides used for fluorescence polarization measurements were synthesized N-terminally labeled with fluorescein and purified by Tufts University Core Services. The buffer used in the fluorescence polarization assay was 20 mM Tris pH7.5, 150 mM NaCl, 1 mM DTT and 0.01% Triton X-100. The protein concentration was varied from 0.97 to 1000 μM and the peptide concentration was 40 nM. An excitation wavelength of 485 nm and an emission wavelength of 528 nm were used. The data were measured at 25°C and corrected for background by subtracting the free-labeled peptide background. The data were collected by the Synergy 2 MAR CCD and corrected for background by subtracting the free-labeled peptide background. The data were measured at 25°C and corrected for background by subtracting the free-labeled peptide background. The data were collected by the Synergy 2 MAR CCD and corrected for background by subtracting the free-labeled peptide background. The data were measured at 25°C and corrected for background by subtracting the free-labeled peptide background. The data were collected by the Synergy 2 MAR CCD and corrected for background by subtracting the free-labeled peptide background. The data were measured at 25°C and corrected for background by subtracting the free-labeled peptide background.

2.3. Structure determination and refinement

Diffraction data were reduced using the HKL2000 suite [17]. Initial molecular replacements (MR) was performed with the program PHASER [18]. The search models were modified by alignment of the target sequence using the FFSAS03 server [19] and SCWRL [20]. For PHF20.83-150, automated model tracing was performed using both BUCCANEER [21]and ARP/wARP [22]. For PHF20.4-69, building of the final model was initiated from the coordinates of PDB [23] entry 2EQM, which were placed into a RESOLVE [24]-modified electron density map by the program MOLREP [25]. For both structures, intermediate model refinement and rebuilding were performed with REFMAC [26] and COOT [27], respectively. Model geometry was monitored on the MOLPROBITY server [28,29]. Model depositions in the PDB were prepared with PDB_EXTRACT [30].

3. Results

3.1. The N-terminal Tud1 domain has an atypical aromatic cage and occluded secondary binding site

The crystal structure of the first Tudor domain of PHF20 (residues 1–81, referred to as Tud1 hereafter) was determined at 2.35 Å by molecular replacement (Table 1). PHF20-Tud1 comprises a typical Tudor domain with five β-sheets arranged antiparallel in the crystallographic asymmetric unit. The resolution limit is determined by the restrained least-squares refinement of the model against the diffraction data. The R-factor and R-free are calculated from the difference electron density map by the program MOLREP [25]. The model was deposited in the PDB (PDB code 3Q1J) with the coordinates and structure factors listed in the PDB. The crystal structure of the second Tudor domain of PHF20 (residues 83–150) was determined at 2.35 Å by molecular replacement (Table 1). PHF20-Tud2 comprises a typical Tudor domain with five β-sheets arranged antiparallel in the crystallographic asymmetric unit. The resolution limit is determined by the restrained least-squares refinement of the model against the diffraction data. The R-factor and R-free are calculated from the difference electron density map by the program MOLREP [25]. The model was deposited in the PDB (PDB code 3QII) with the coordinates and structure factors listed in the PDB.

Table 1

<table>
<thead>
<tr>
<th>Data collection and refinement statistics.</th>
<th>PHF20 (residues 1–81)</th>
<th>PHF20 (residues 83–150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code</td>
<td>3Q1J</td>
<td>3QII</td>
</tr>
<tr>
<td>X-ray source</td>
<td>Rigaku FR-E</td>
<td>CLS beamline</td>
</tr>
<tr>
<td>X-ray wavelength (Å)</td>
<td>1.54Å</td>
<td>4.0Å</td>
</tr>
<tr>
<td>Detector</td>
<td>Rigaku R-Axis 4-HTC</td>
<td>MAR 300 CCD</td>
</tr>
<tr>
<td>Oscillation range (Å)</td>
<td>360 × 0.5°</td>
<td>180 × 1°</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>P4₂,2,2</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>60.50, 41.06, 28.63</td>
<td>48.88, 48.88, 96.27</td>
</tr>
<tr>
<td>x, y, z (°)</td>
<td>90, 110.33, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (outer shell, Å)</td>
<td>20.00–2.35</td>
<td>50.00–2.30</td>
</tr>
<tr>
<td>Unique observed HKLs</td>
<td>2828 (288)</td>
<td>5600 (237)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (100.0)</td>
<td>99.2 (86.8)</td>
</tr>
<tr>
<td>Friedel Redundancy</td>
<td>3.4 (3.3)</td>
<td>12.3 (6.4)</td>
</tr>
<tr>
<td>Rsym</td>
<td>1.00 (0.654)</td>
<td>0.069 (0.318)</td>
</tr>
<tr>
<td>l&gt;θ (°)</td>
<td>18.4 (2.2)</td>
<td>46.3 (3.6)</td>
</tr>
<tr>
<td>MR model PDB code</td>
<td>2RHU</td>
<td>204X</td>
</tr>
<tr>
<td>Post-MR density modification</td>
<td>DM, RESOLVE</td>
<td>DM</td>
</tr>
<tr>
<td>Final model refinement</td>
<td>AUTOBUSTER</td>
<td>REFMAC [26]</td>
</tr>
<tr>
<td>Refinement resolution (Å)</td>
<td>19.08–2.35</td>
<td>30.00–2.30</td>
</tr>
<tr>
<td>Number of reflections (of which free set)</td>
<td>2804 (131)</td>
<td>5533 (266)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.205/0.260</td>
<td>0.235/0.268</td>
</tr>
<tr>
<td>Number of atoms/mean B factor (Å²)</td>
<td>557/43.5</td>
<td>431/46.2</td>
</tr>
<tr>
<td>Protein</td>
<td>552/43.5</td>
<td>421/46.3</td>
</tr>
<tr>
<td>Water</td>
<td>5/41.1</td>
<td>10/42.3</td>
</tr>
<tr>
<td>RMSD bond lengths (Å)/angles (°)</td>
<td>0.010/1.0</td>
<td>0.016/1.4</td>
</tr>
<tr>
<td>PROCHECK</td>
<td>Ramachandran residues in most favored regions/total</td>
<td>46/47</td>
</tr>
<tr>
<td>Additional allowed</td>
<td>5/56</td>
<td>1/47</td>
</tr>
</tbody>
</table>
barrel form (Fig. 2A). Strong electron density was apparent for residues 4–69. Terminal residues were disordered and without supporting electron density. As a result, they were omitted in this model. Analysis of the structure using the PISA [31] server did not indicate the presence of interfaces that might lead to homo-oligomerization in solution. A search of the PDB database using the protein structure comparison service Fold at the European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>) showed the domain to closely resemble other Tudor domains including the Tud2 structure reported herein (rmsd = 1.5 Å), human SMN protein (PDB 1MHN, rmsd = 1.32 Å), and human Tudor domain-containing protein 3 (PDB 3PMT, rmsd = 1.68 Å) [32]. Good alignment was also observed with other members of the Tudor family, including the individual Tudor domains of 53BP1 (PDB 2IG0) [33], and JMJD2A (PDB 2QQS) [34]).

The N-terminus of Tud1 occupies a binding site opposite to the putative aromatic cage. Binding of the N-terminal tail is stabilized by three primary interactions (Fig. 2). Firstly, interaction between the pyrrolidine of Pro5 and the indole of Trp60 stabilizes the far N-terminus of the tail. Hydrogen bond interactions between the side chain of R8 and the mainchain atoms of Trp60, Ser62 and Leu65 serve as a second anchoring point along the tail. Finally, hydrophobic interactions between the side chain of F13 and a pocket formed by Ile33, Val43, Trp60, and Leu65 provide a significant surface for N-terminal stabilization. This results in a well-ordered N-terminal tail, with the exception of residues 1–3.

PHF20-Tud1 possesses an atypical, but complete aromatic cage comprising Asp23, Tyr29, Phe47, Trp50 and Tyr54 (Fig. 2). Comparison of this site with the aromatic cages of other structurally characterized Tudor domains reveals the backbone atoms of Trp50 and Tyr54 to be significantly displaced relative to equivalent residues of other competent aromatic cages. This distortion appears to result from the presence of a 310-helix turn spanning residues Asn51 to Asp55, which shortens strand β4 and generates a novel Tudor domain variant. Not seen in other Tudor domains of known structure, the he 310-helix results in occlusion of a potential aromatic cage by the positioning of Trp50 toward the center of the binding site, preventing recognition of methylated residues by this domain.

3.2. Tud2 domain of PHF20 displays canonical Tudor architecture

The structure of residues 83 to 150 (PHF20-Tud2) also comprises a canonical Tudor domain with 5 anti-parallel β-strands. The structure was determined by molecular replacement to 2.3 Å resolution (Table 1, Fig. 3A). Electron density was apparent for all residues in this expression construct. PISA analysis also failed to identify potential dimerization interfaces amongst the crystal contacts. Structural alignment of the Tud2 domain with other Tudor domains reveals strong structural similarity with the aforementioned Tud1 of PHF20, as well as other members of the Tudor domain family with root mean squared deviations ranging from 1.3–1.9 Å on alignment of the backbone atoms.

The aromatic cage of PHF20-Tud2 comprises Trp97, Tyr103, Phe120, Asp122, and Val124. Alignment of PHF20-Tud2 with other individual Tudor domains from proteins such as 53BP1, PHF1, PHF19, and TDRD3 reveals excellent structural conservation for the residues comprising the aromatic cage. The alignment of this domain with 53BP1 is presented in Fig. 3B to illustrate the conservation of the methyllysine binding site (Fig. 3B). By amino acid
sequence, the cages of PHF20 is most similar to that of Lamin B receptor, which comprises Trp, Tyr, Tyr, Asp, and Thr52 (Fig. 3C). Alignment of the PHF20-Tud2 domain with the NMR structures of Lamin B (PDB ID 2L8D and 2DIG), however, reveals little structural difference in the aromatic cage region, save for the rotation of PHF20-Trp97 90° relative to the Lamin B structures. The structure reported herein reveals dimerization of the Tud2 domain mediated by two intermolecular disulfide bonds from two symmetry-related molecules (Fig. 3D). The two disulfide bonds are formed between Cys96 of molecule A and Cys100 of molecule B, and Cys100 of molecule A and Cys96 of molecule B. The dimeric form of PHF20-Tud2 was not apparent on size exclusion chromatography and no additional evidence supports the formation of dimeric PHF20-Tud2 in solution or in vivo.

3.3. The tandem Tudor domains of PHF20 are highly similar to those of PHF20L1

The tandem Tudor domains of PHF20 are highly similar to those of the PHF20L1 protein with a large number of strictly conserved residues evident in the sequence alignment for this region (Fig. 4A). Comparison of PHF20-Tud1 structure with those of first Tudor domain of PHF20L1 (PHF20L1-Tud1) (Fig. 4B, PDBs 2JTP and 2EQM, NMR structures) reveals a similar potential peptide binding site, occluded by the N-terminal tail, is present in the PHF20L1-Tud1 (Fig. 4B and C). All residues of the N-terminal tail have been described for the 2EQM structure and the backbone atoms of this peptide correspond well with those of PHF20-Tud1 (Fig. 4B and C). The 2EQM entry also describes an insertion of the
F13 into the hydrophobic pocket of the peptide binding cleft (Fig. 4c). Support for the importance of interactions between the proline and arginine in ordering the N-terminal peptide in this cleft is provided by the 2JTP structure in which residues 1 to 6 have been omitted. The N-terminal peptide of 2JTP is not ordered and the F13 residue is no longer anchored into the hydrophobic pocket, assuming numerous conformations in the 15 reported NMR models (Fig. 4C and D).

All residues of the putative PHF20-Tud1 aromatic cage are conserved in the corresponding cage of PHF20L1 and, in likewise fashion, the potential binding pockets are occluded by insertion of a Trp residue (Fig. 4B). However, the independent NMR structures of this domain (PDBs 2JTP and 2EQM) reveal differences between each other and with respect to PHF20-Tud1. Neither structure of PHF20L1 presents a pseudo-helical insertion between residues 51 and 55 in any of the deposited NMR models (Fig. 4B). Additionally, the position of Tyr54 differs dramatically between the two PHF20L1 structures. In 2JTP, Tyr54 is completely flipped out of the pocket and highly flexible when compared amongst the models used to generate the average deposited model (Fig. 4D). In the 2EQM structure, however, Tyr54 is structurally conserved with respect to PHF20-Tud1 and less flexibility is apparent among the structure ensemble, supporting the PHF20-Tud2 structure of the occluded aromatic cage.
Comparison of the PHF20-Tud2 domain with the second Tudor domain of PHF20L1 (PHF20L1-Tud2, PDB 2EQU) reveals little difference between the two proteins (Fig. 4E). Sequence-wise, the aromatic cage is conserved between the two proteins (Fig. 4A). The only apparent structural difference at the methyllsine binding site is the rotation of the Trp97 sidechain in the PHF20L1-Tud2 structure (Fig. 4E).

3.4. Tud2 but not Tud1 binds dimethylated histone substrates

Potential interactions between the Tudor domains of PHF20 and various histone substrates were probed using fluorescence anisotropy measurements with fluorescence labeled ligands. Tud1 of PHF20 failed to exhibit any interactions with the peptides screened. Tud2 exhibited a preference for dimethylated substrates (Fig. 5A). No significant affinity was measurable for the other methylation states of this peptide, showing a clear preference of the Tud2 pocket for dimethyllysine (Fig. 5B).

4. Discussion

Methylation of histone tails influences overall chromatin structure and the accessibility of DNA segments. Recognition of these methyl marks has been attributed to the Royal Family of proteins, which includes the Tudor domain subfamily. Tudor-domain-containing proteins may possess one or more copies of the domain. tandem Tudor domains (TTD), in which two individual domains reside in very close proximity to one another, assume a wide variety of architectures (see Supplemental Fig. 1). The diversity of TTD arrangements is likely due to the variations of interdomain linkers, as well as the inherent flexibility of the Tudor domain itself. We have divided the TTD of PHF20 into two separate domains to improve the stability of our expression constructs for structural studies as the TTD was not stable in solution. Both structures, however, suggested a mechanism for peptide recognition by an extended binding site on the face opposite the occluded aromatic cage. The interaction between the N-terminal tail and this binding cleft is similar to that observed between the first pocket of L3MBTL1 and the C-terminal tail of a symmetry-related molecule where the “REPSSA” peptide segment forms a type II β turn with the pyrrolidine ring residing within the hydrophobic pocket. Both structural features were observed in crystallo-graphic and NMR structures, suggesting a function relevance in both crystal and solution states.

The PHF20-Tud2 domain was shown here to have a preference for dimethylated lysine residues. Structural comparison with PHF20L1 suggests that the homolog is also capable of methyllsine recognition. In contrast, the Tudor domain of Lamin B was recently reported to not bind these modified peptides, demonstrating that slight differences in the aromatic cages of Tudor domain family members have significant impact on peptide recognition.

The PHF20 protein has been isolated from MLL methyltransferase complexes involved in the deposition of methyl marks along histone tails. The dimethylation of H3K4 is carried out by the SET family of enzymes, which comprises MLL1 through 5 and SET1A/B. Studies of the localization of H3K4me2 across the whole genome have indicated that this mark is associated with genes poised for transcription, while the trimethylation of H3K4 is associated with actively transcribed genes. The ability of PHF20-Tud2 to recognize dimethylated histone substrates may be an important contributor to the regulation and/or assembly of MLL complexes.

Following release of the crystal structures described here, additional crystal structures representing Tud1 [42] and Tud2 [43] respectively, were also released to the public. Those models can be aligned with rmsds of 0.47 Å (0.46 Å in case of chain B of entry 3SD4) and 0.49 Å (0.29 Å in case of chain B of entry 3P8D) against their respective equivalents 3Q1J and 3QII, which are described here.

Acknowledgments

We would like to thank Amy K. Wernimont, Chao Xu, Peter Loppnau and Lissete Crombet for advice and technical assistance. This research was supported by the Structural Genomics Consortium, a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research (CIHR), the Canadian Institutes for Health Research (CIHR) and the Canadian Institutes for Health Research (CIHR).
Appendix A. Supplementary data


References


