Crystal Structure of the C-Terminal Domain of Human DPY-30-Like Protein: A Component of the Histone Methyltransferase Complex

Xianping Wang1,2, Zhiyong Lou2,3, Xiuhua Dong1, Wen Yang4, Yong Peng1, Bin Yin1, Yanhua Gong1, Jiangang Yuan1, Weihong Zhou4, Mark Bartlam4⁎, Xiaozhong Peng1⁎ and Zihe Rao2,3,4

1National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100005, China
2Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China
3National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
4College of Life Sciences and Tianjin State Laboratory of Protein Sciences, Nankai University, Tianjin 300071, China

Received 15 March 2009; received in revised form 18 May 2009; accepted 20 May 2009
Available online 27 May 2009

Edited by M. Guss

The conserved DPY-30 is an essential component of the dosage compensation complex that balances the X-linked gene expression by regulation of the complex formation in Caenorhabditis elegans. The human DPY-30-like protein (DPY-30L) homolog is a conserved member of certain histone methyltransferase (HMT) complexes. In the human MLL1 (mixed-lineage leukemia-1) HMT complex, DPY-30L binds to the BRE2 homolog ASH2L in order to regulate histone 3–lysine 4 trimethylation. We have determined the 1.2-Å crystal structure of the human DPY-30L C-terminal domain (DPY-30LC). The DPY-30LC structure, harboring the conserved DPY-30 motif, is composed of two α-helices linked by a sharp loop and forms a typical X-type four-helix bundle required for dimer formation. DPY-30LC dimer formation is largely mediated by an extensive hydrophobic interface with some additional polar interactions. The oligomerization of DPY-30LC in solution, together with its reported binding to ASH2L, leads us to propose that the hydrophobic surface of the dimer may provide a platform for interaction with ASH2L in the MLL1 HMT complex.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: DPY-30L; crystal structure; dimer; HMT complex; X-type four helix bundle

*Corresponding authors. M. Bartlam is to be contacted at College of Life Sciences and Tianjin State Laboratory of Protein Sciences, Nankai University, Tianjin 300071, China; X. Peng, National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100005, China. E-mail addresses: bartlam@nankai.edu.cn; pengxiaozhong@pumc.edu.cn.

Abbreviations used: AKAP, A-kinase anchoring protein; D/D, docking and dimerization; DPY-30L, DPY-30-like; DPY-30LC, DPY-30L C-terminal domain; HMT, histone methyltransferase; HNF-1α, hepatocyte nuclear factor 1α; PKA, protein kinase A; SAD, single-wavelength anomalous dispersion; Se-Met, selenomethionyl.

0222-2836/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
Introduction

DPY-30 and its homologs are mainly expressed in the cell nucleus and are conserved components in histone methyltransferase (HMT) complexes in yeast, Caenorhabditis elegans and humans. In these complexes, the DPY-30-like (DPY-30L) protein can directly interact with ASH2L and is likely to form part of a four-subunit subcomplex, alongside other proteins (WDR5 and RBBP5), of the larger SET-domain-containing HMT complex. The large complex regulates histone 3 lysine 4 (H3K4) methylation and activates target gene transcription. 

Li et al. reported for C. elegans the first interaction between DPY-30 and the ASH2L homolog Y17G7B.2 in a genome-wide yeast two-hybrid screen. DPY-30 is therefore not only a component of the dosage compensation complex but also an important member of the histone methylation complex in C. elegans. Furthermore, as Schizosaccharomyces pombe RBP2 homolog LID2 (tri- and di-demethylases of H3K4) co-purifies with BRE2 and SDC1 of the yeast COMPASS H3K4 methyltransferase complex members, raising the possibility that the HMTs and demethylases with specificity for the same histone mark have common binding partners, such as SDC1 and BRE2.

The human DPY-30L protein, a homolog of C. elegans DPY-30, has been identified as a component of HMT complexes. Epigenetic control is a central molecular mechanism of dosage compensation, and histone methylation is important for regulating the transcriptional accessibility of chromatin. Differences in histone lysine methylation may lead to opposite roles in gene transcription, and methylation of H3K4 usually results in the positive regulation of gene transcription. A number of large HMT complexes are conservative and respond to the monomethylation, dimethylation and trimethylation of H3K4. Examples include the trithorax group in Drosophila, the SET1 or COMPASS complex in Saccharomyces cerevisiae and MLL1 (mixed-lineage leukemia-1) in humans. The yeast SDC1 protein, a homolog of DPY-30 in C. elegans, can interact with BRE2 directly and form the SET1 complex with other members—SWD1, SPP1, SWD2, SWD3 and SHG1. They influence the SET1 HMT activity for H3K4 tails rather than the presence of SET1 in specific chromosomal regions. The absence of SDC1 and BRE2 reduces the dimethylation of H3K4 and severely affects trimethylation.

DPY-30L is a nuclear protein universally expressed in human tissue. It includes a short DPY-30 motif at its C-terminal that is highly conserved in humans, mice, Drosophila and C. elegans (Fig. 1). The DPY-30 motif also exists in other important proteins, including ankyrin repeats, adenylyl kinase and nucleoside diphosphate kinase homolog 5. However, to the best of our knowledge, no three-dimensional structure of the DPY-30 motif has been determined to date. In this study, we report the first crystal structure of a DPY-30 motif from the human DPY-30L protein and that it exists as a homodimer. One DPY-30L monomer consists of two short and flexible terminal helices linked by a short turn. These short terminal helices contribute toward dimer formation through a typical X-type four-helix bundle involving strong hydrophobic interactions and hydrogen bonds between the termini of each monomer.

Results

The monomer fold of the DPY-30 motif

The DPY-30 motif was initially identified from the DPY-30 protein in C. elegans. It is about 40 residues in length and is proposed to form two α-helices. This motif is found in a wide variety of contexts, including the human DPY-30L protein, which is a component of HMT complexes. The disordered N-terminal region of full-length human DPY-30L makes crystallization difficult. Therefore, we selected the 55 C-terminal amino acids of human DPY-30L (termed DPY-30LC hereinafter) for crystallization and subsequent structure determination. The crystal structure of DPY-30LC, including residues 45–99 of DPY-30L, was determined to date. The absence of SDC1 and BRE2 reduces the dimethylation of H3K4 and severely affects trimethylation.

The yeast SDC1 protein, a homolog of DPY-30 in C. elegans, can interact with BRE2 directly and form the SET1 complex with other members—SWD1, SPP1, SWD2, SWD3 and SHG1. They influence the SET1 HMT activity for H3K4 tails rather than the presence of SET1 in specific chromosomal regions. The absence of SDC1 and BRE2 reduces the dimethylation of H3K4 and severely affects trimethylation.

In the human MLL1 complex, DPY-30L can directly interact with the BRE2 homolog ASH2L and plays a role similar to that of the SDC1 and BRE2 interaction in the yeast Set1 complex. DPY-30L is a nuclear protein universally expressed in human tissue. It includes a short DPY-30 motif at its C-terminal that is highly conserved in humans, mice, Drosophila and C. elegans (Fig. 1). The DPY-30 motif also exists in other important proteins, including ankyrin repeats, adenylyl kinase and nucleoside diphosphate kinase homolog 5. However, to the best of our knowledge, no three-dimensional structure of the DPY-30 motif has been determined to date. In this study, we report the first crystal structure of a DPY-30 motif from the human DPY-30L protein and that it exists as a homodimer. One DPY-30L monomer consists of two short and flexible terminal helices linked by a short turn. These short terminal helices contribute toward dimer formation through a typical X-type four-helix bundle involving strong hydrophobic interactions and hydrogen bonds between the termini of each monomer.
by the single-wavelength anomalous dispersion (SAD) phasing method and refined to 1.2-Å resolution with a final $R$-factor of 19.7% and an $R_{	ext{free}}$ of 22.7% (Fig. 2a and Table 1). From good electron density (Fig. 2b), the final model contains 198 amino acid residues in four independent polypeptides (chains A–D), together with 378 ordered water molecules, one 1,6-hexanediol molecule and two dithiothreitol molecules. The N-terminal amino acid and two C-terminal amino acids are absent from the electron density map for chain A. Each independent polypeptide contains two $\alpha$-helices and a sharp reversal linker (Fig. 2a). Based on the molecular weight of the protein estimated from gel filtration and analytical centrifugation (Fig. 3), the tetramer in the asymmetric unit is expected to be a crystal packing artifact. Therefore, we analyzed the DPY-30L$_C$ structure as a dimer. The two protomers form a canonical X-type four-helix bundle with an antiparallel arrangement, as illustrated in Fig. 2c and d. The surface of DPY-30L$_C$ is compactly packed and forms a large hydrophobic surface in the dimer interfaces (Fig. 4).

Alignment of the protomers by least-squares superposition of main-chain atoms (residues 50–95) yields a root-mean-square deviation (r.m.s.d.) of 0.4 Å, indicating that protomers adopt essentially identical conformations. Two protomers form a compact dimer via predominantly hydrophobic interactions. The core region of the protomer is formed by amino acids from Thr53 to Phe95. The residues from Thr53 to Glu75 make up the first $\alpha$-helix, divided into $\alpha_1A$ and $\alpha_1$ by a 66.3° bend about

![Fig. 2. Structure of human DPY-30L$_C$. (a) Crystal structure of the human DPY-30L$_C$ monomer. The structure is shown colored from blue at the N-terminus to red at the C-terminus. Secondary structure elements are labeled. (b) Representative electron density covering part of the DPY-30L$_C$ dimer interface. The monomers are shown in gold and blue stick representation. The $2F_o - F_c$ electron density map is shown as a gray mesh and contoured at 2.0$\sigma$. (c and d) The DPY-30L$_C$ dimer interface showing the hydrophobic interactions between helix 1 and helix 2, respectively. The monomers are shown in gold and blue ribbon representation, with hydrophobic side chains shown in stick representation. The views in (c) and (d) are related by a rotation of 180°.](image-url)
Pro63. The second helix includes residues from Pro80 to Phe95, with a short turn (Arg76–Asn79) linking the two helices. Helices 1 and 2 in a single monomer are almost orthogonal to one another, with an interhelical angle of 115°. The 11 major residues participate in the intermolecular hydrophobic surface of the dimer. In total, 30% of the available surface area in the monomer is buried in the interface upon dimer formation. The surface area buried in the interface of the dimer is ∼1900 Å², or ∼950 Å² per monomer.

Dimer architecture and interaction between protomers

The α-helices forming the antiparallel bundle in the dimer interface are mainly held together by extensive hydrophobic and van der Waals interactions between protomers. The leucine and isoleucine side chains form a cluster within the hydrophobic core, including Leu57, Val62, Leu65, Leu66, Met69 (Leu69 in the wild-type protein), Val71 and Leu72 on helix 1 as well as Ile81, Leu84, Leu88 and Leu89 on helix 2 of each monomer (Fig. 2c and d). Four hydrogen bonds also assist in dimer formation. The main-chain oxygen atoms of Leu48 and Leu51 of one protomer interact with the side chain of Arg76 of the neighboring protomer in the dimer. The hydrophobic residues Leu72 and Leu84 are buried in the interface of the dimer, and the hydrophobic surface of the dimer is composed of Leu48, Leu51, Leu57, Val62, Leu65, Leu66, Met69 and Val71 of each monomer. Pro63 is located in the α-helix bend, where it can interact with several amino acids, including Leu65, Leu66 and Gln67. With Pro63 in this specific location, the following helix 1 can bend into the hydrophobic center and enable closer interaction with helix 1 from a neighboring protomer, which benefits dimer formation.

Comparison of the DPY-30 motif with other X-type four-helix bundles

The three-dimensional structure of a DPY-30 motif has not been reported previously, but the motif is highly conserved in amino acid sequence alignment among C. elegans, Drosophila, mice and humans (Fig. 1). The crystal structure of DPY-30LC reveals that the DPY-30 motif forms a canonical X-type bundle structure. Superposition of the DPY-30LC X-type bundle with related structures, including the protein kinase A (PKA) regulatory subunit fragments RIIα (r.m.s.d. of 2.1 Å for 37 Cα atoms) and RIIβ (r.m.s.d. of 2.4 Å for 30 Cα atoms), the calcium-binding protein S100B (r.m.s.d. of 3.3 Å for 35 Cα atoms) and the transcriptional activator HNF-1α (hepatocyte nuclear factor 1α; r.m.s.d. of 2.7 Å for 39 Cα atoms), indicates that all proteins form a helical bundle via hydrophobic interactions despite sharing low sequence similarity (Fig. 4a).

Previous studies have shown that the PKA regulatory subunit fragment RIIα [amino acids 1–44; Protein Data Bank (PDB) code 1R2A] is relatively close to DPY-30L in terms of secondary structure (Fig. 4b and c). A structure-based sequence alignment shows some conservation in the residues that form the hydrophobic dimerization interface, particularly those in helix α2, while the N- and C-termini are less conserved (Fig. 4a). The PKA RIIα structure can be subdivided into two functional regions, the first of which encompasses residues responsible for dimerization and formation of the protein-binding surface. The so-called docking and dimerization (D/D) domain of RIIα consists of two α-helices that form an X-type bundle of four helices during dimerization. If helix 1 alone (residues 61–75) is superimposed with helix 1 of RIIα (residues 11–25), the least-squares fits between Cα atoms are 0.74 Å (chain A) and 0.77 Å (chain C), respectively. In contrast, superposition of helix 2 of RIIα (residues 30–43) yields respective r.m.s.d. values of 1.23 and 1.10 Å for chains A and C. This implies that helix 2 of the DPY-30LC motif is structurally less conservative than helix 1. The DPY-30LC N-terminal domain (residues 46–60) is substantially different from RIIα (residues 1–11), with an r.m.s.d. of 3.2 Å. The N-terminal domain of DPY-30L has a bend of about 66° and is oriented toward the dimer interface such that its alkaline end protrudes into the hydrophobic platform. This would suggest that the N-terminal
domain is a key contributing factor toward the recognition of target binding proteins by DPY-30L.

The RIIα D/D domain can provide an extended hydrophobic and solvent-accessible groove as an interface for protein and peptide interactions (Fig. 4c and e). The residues Leu9 and Leu13 (helix 1) as well as Leu9′ and Leu13′ (helix 2) located in the hydrophobic groove can interact directly with the A-kinase anchoring protein (AKAP) peptide Ht31 and AKAP79 with slight adjustments. The site is also highly conserved in the DPY-30 motif and is located in the hydrophobic surface. It is speculated that the residues forming this site are important for recognition of DPY-30L-binding proteins.

The N-terminal hydrophobic residues (5–8) of the RIIα D/D domain are AKAP-binding determinants, although they are required to undergo a slight structural change. The DPY-30C N-terminal domain (residues 46–60) differs from the RIIα D/D N-terminal domain, which contains several polar residues, such as Asp4, Glu5 and Arg10. Furthermore, unlike the RIIα structure (Fig. 4e), the electrostatic potential of DPY-30C shows that the dimer has a cleft in the hydrophobic surface with a positive protrusion (Fig. 4d) that may provide more restraint on the permitted binding partners. Electrostatic properties are also known to play a role in the structure and stability of the RIIα D/D domain, influencing, for instance, quaternary packing. The differences in the electrostatic properties of the N-terminal residues in DPY-30C may therefore provide clues for the specific function of DPY-30C and its particular binding partners.

**Discussion**

The crystal structure of DPY-30C shows that its DPY-30 motif exists as a compact dimer and forms a typical X-type four-helix bundle in an antiparallel arrangement. The dimer interface is predominantly hydrophobic and is composed by long-side-chained amino acids. We propose that DPY-30C might interact with ASH2L via this hydrophobic surface. The putative interaction of the DPY-30 motif with ASH2L can be inferred from the different structures of RIIα and AKAP complexes (Fig. 4c and e). The long-side-chained hydrophobic residues of RIIα form an extended hydrophobic groove for AKAP peptide binding, and about 1370 Å², or about 25% of the total surface area (5350 Å²) of the D/D domain, becomes solvent inaccessible upon binding either of
the AKAP peptides. The similarity between the crystal structures of the two proteins implies a hypothetical mechanism for the binding of DPY-30Lc to ASH2L via the hydrophobic surface of the dimer.

Several reports point toward the emergence of the X-type four-helix bundle as an adaptable yet selective domain that provides a platform for protein–protein interactions. X-type helical bundles are usually composed of two α-helices that form a dimer in an antiparallel arrangement and play roles as signal transduction factors. For example, the X-type four-helix bundle of HNF-1α can bind to the DcoH dimer via an exposed hydrophobic patch. S100B, another X-type four-helix bundle, forms a dimer for the interaction with p53 once Ca2+ binds, exposing a surface formed by the four-helix bundle and EF hand to recruit two p53 molecules per dimer.

PKA RIIα and RIIε both form typical X-type four-helix bundles and dimers to form an extensive hydrophobic groove for binding AKAP peptides. A crystal structure of the helical motif from D-AKAP2 bound to PKA RIIα (PDB code 2HWN) has revealed the mechanism for peptide binding, showing that the AKAP helix binds into a stable hydrophobic groove formed by helix 1 of each protomer in the interface of the four-helix bundle. Comparison of the DPY-30Lc structure reported here with the peptide-bound RIIα structure offers some clues for specificity of the peptide recognition. The groove formed by the symmetrical DPY-30Lc dimer provides a less flat, less hydrophobic binding surface than the equivalent peptide-binding groove of RIIε. A pronounced depression in the center of the DPY-30Lc dimer surface is formed by Leu65, Leu88 and Leu89 and
flanked on either side by Leu66. More significantly, the bend in helix 1 of DPY-30LC results in a narrower hydrophobic peptide-binding groove. In particular, Arg54 of each protomer is directed in toward the binding groove and should place more stringent requirements on the binding specificity of the DPY-30LC dimer. However, as the RIfc D/D domain undergoes a dynamic ordering of the N-terminus in one protomer upon AKAP peptide binding, it is unclear whether DPY-30LC is required to undergo a similar reordering or conformational change when interacting with its binding partners.

Our results presented here reveal, for the first time, the three-dimensional structure of a DPY-30 motif from human DPY-30L. The compact dimer formed by the C-terminal domain of DPY-30L differs from previously reported structures of X-type four-helix bundles, both in terms of the surface properties and their polar termini. The crystal structure of DPY-30LC is consistent with reports that the X-type four-helix bundle provides a platform for protein–protein interactions. The less hydrophobic surface may provide a potential binding region for ASH2L, offering a significant starting point for further binding studies.

Materials and Methods

Cloning, expression, purification and crystallization

Human DPY-30LC (residues 45–99) was cloned into the pGEX-6p-1 (GE Healthcare) vector and transformed in *Escherichia coli* BL21(DE3). When the optical density at 600 nm reached 0.6–0.8 in LB medium at 37 °C, it was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside and continuously grown for an additional 10 h at 16 °C. Then, bacterial cells were sonicated in phosphate-buffered saline buffer and centrifuged at 15,000 rpm for 30 min at 4 °C and the supernatant was harvested. The first purification step involved a glutathione affinity column (GE Healthcare) and washing the unbound protein in 20 mM phosphate-buffered saline buffer and centrifuged at 15,000 rpm, pH 7.5, at 20 °C at 60,000 rpm. A set of 200 scans was collected at 1-min intervals at 280 nm. Analysis of the continuous sedimentation coefficient distribution was performed with the software Sedfit v. 0.93.25 Size-exclusion chromatography was performed with an AKTA Explorer (GE Healthcare). For chromatographic separations, a Superdex 75 6 HR 10/300 column (GE Healthcare) was equilibrated with 20 mM Hepes and 100 mM NaCl, pH 7.5, at an eluent flow rate of 0.5 ml/min. For calibration of the gel-filtration column, the Superdex 75 column was balanced with a buffer containing 20 mM Tris–HCl, pH 8.0, and 100 mM NaCl. Four proteins, aprotinin (6.512 kDa), ribonuclease A (13.7 kDa), albumin (67 kDa), and bovine serum albumin (67 kDa), were used for molecular mass standards and solved in the same buffer as mentioned above. The originPro 7.5 software was used for the calibration curve. DPY-30LC was calibrated with the curve.

The results of data analysis showed sediment of DPY-30LC at 1.47S (Fig. 3a), corresponding to a molecular mass of 15.4 kDa (Fig. 3b). This in turn corresponds to the mass of the dimer of DPY-30LC, which is in good agreement with the results of gel-filtration chromatography (Fig. 3c).

Accession number

Coordinates and structure factors have been deposited in the PDB with accession number 3G36.

Acknowledgements

This work was supported by grants from the National Natural Sciences Foundation of China (nos. 30430200, 30721063) and the Ministry of Science.

†http://www.pymol.org
and Technology Project “973” (no. 2007CB914301) and Project “863” (nos. 2006AA02Z13 and 2006AA02A304). We thank Yu Guo for help in data collection and Xiaohong Zhou and Dr. Hui Wang for valuable discussions. We also thank Xiaoxia Yu for analysis of the analytical ultracentrifugation data.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.05.061

References