Structure of the Conserved Core of the Yeast Dot1p, a Nucleosomal Histone H3 Lysine 79 Methyltransferase*

Ken Sawada‡§, Zhe Yang‡§, John R. Horton‡, Robert E. Collins‡§, Xing Zhang‡, and Xiaodong Cheng‡

From the ‡Department of Biochemistry, §Graduate Program in Biochemistry, Cell, and Development Biology, Emory University School of Medicine, Atlanta, Georgia 30322

Received for publication, May 27, 2004, and in revised form, July 21, 2004 Published, JBC Papers in Press, July 29, 2004, DOI 10.1074/jbc.M405902200

Methylation of Lys79 on histone H3 by Dot1p is important for gene silencing. The elongated structure of the conserved core of yeast Dot1p contains an N-terminal helical domain and a seven-stranded catalytic domain that harbors the binding site for the methyl-donor and an active site pocket sided with conserved hydrophobic residues. The S-adenosyl-L-homocysteine exhibits an extended conformation distinct from the folded conformation observed in structures of SET domain histone lysine methyltransferases. A catalytic asparagine (Asn479), located at the bottom of the active site pocket, suggests a mechanism similar to that employed for amino methylation in DNA and protein glutamine methylation. The acidic, concave cleft between the two domains contains two basic residue binding pockets that could accommodate the outwardly protruding basic side chains around Lys79 of histone H3 on the disk-like nucleosome surface. Biochemical studies suggest that recombinant Dot1 proteins are active on recombinant nucleosomes, free of any modifications.

Histones can be modified in many ways to affect gene expression, including acetylation, phosphorylation, ubiquitination, methylation (reviewed in Refs. 1 and 2), and sumoylation (3). Evidence accumulated over the past few years suggests that such modifications constitute a “histone code” that directs a variety of processes involving chromatin (4, 5). There are currently many known sites of lysine methylation on histones and additional sites of modification are still being uncovered. Methylation at these sites, in combination with other modifications (or demodifications) at nearby residues, generates a complex spectrum of modifications at Lys79 of histone H3 (12). Methylation of H3 Lys79 is important for gene silencing and the proper localization of the SIR (silent information regulator) complex in S. cerevisiae (8, 11). A homology and sequence analysis (13) suggested that Dot1p possesses S-adenosyl-L-methionine (AdoMet)1 binding motifs characteristic of class I methyltransferases (MTases) (14), similar to the ones in protein-arginine MTases (PRMTs) that modify arginines on many proteins including histones H3 and H4. Class I MTases such as Dot1p are distinct from, and do not contain the SET domain, a conserved domain found in HKMTs that methylate lysines 4, 9, 27, or 36 of histone H3 and Lys20 of H4. Thus, entirely different structural scaffolding and unrelated local active site spatial arrangements can catalyze AdoMet-dependent methyl transfer to a protein Lys side chain.

Besides the differences in protein sequence and structure, Dot1p has several unique biochemical properties that distinguish it from the SET domain HKMTs. (i) Dot1p, and its human homolog Dot1L, methylate only nucleosomal substrates, but not free histone H3 protein. Several other enzymes whose targets lie within the highly structured histone core require nucleosome substrates for efficient methylation; this includes Set2, which targets Lys36 of histone H3 (15) and Set8, which methylates Lys36 of histone H4 (16, 17). (ii) A distribution of unmodified, mono-, di-, and trimethylated H3 Lys79 exist simultaneously in yeast (8). Whereas the functions of different methylation states at H3 Lys79 are not fully defined, a complex spectrum of methylation is also found for certain lysine residues methylated by SET domain HKMTs. The S. cerevisiae SET1 protein can catalyze di- and trimethylation of H3 Lys4, and trimethylation of Lys4 is thought to be present exclusively in active genes (18). Human SET7/9 protein, on the other hand, generates exclusively monomethyl Lys4 of H3 (19, 20). Furthermore, DIM-5 of Neurospora crassa generates primarily trimethyl Lys9 of H3, which marks chromatin regions for DNA methylation (19, 21). (iii) Methylation of Lys79 of H3 in S. cerevisiae requires ubiquitination of histone H2B (11, 22). This situation bears similarity to the methylation of Lys6 of H3 by SET1 (23, 24). It was suggested that a ubiquitinated histone H2B might serve as a spacer between adjacent nucleosomes to allow access by enzymes (24).

Here we present the conserved core structure of yeast Dot1p, the founding member of the Dot1p family, in complex with the methyl transfer reaction product S-adenosyl-L-methionine (AdoHcy), at a resolution of 2.19 Å. The structure reveals an

1 The abbreviations used are: AdoMet, S-adenosyl-L-methionine; MTase, methyltransferase; PRMT, protein-arginine MTase; AdoHcy, S-adenosyl-L-homocysteine; HKMT, histone-lysine MTase; CHES, 2-(cyclohexylamino)ethanesulfonic acid; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxyethyl)propane-1,3-diol.
Yeast Dot1p Structure

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The S. cerevisiae Dot1 gene was subcloned from the original pFVL39 construct (8) by PCR into a modified pET28b (Novagen) vector, which adds a short N-terminal MOH-HHHHHH tag and accepts a Ndel-EcoRI insert. Deletions (Δ157 and Δ179) and point mutations were constructed by PCR using the same vector. Cultures of Escherichia coli strain BL21(DE3) Codon plus RIL (Stratagene) containing Dot1 constructs were grown in LB medium at 37 °C for 30 min to an OD600 of ~0.6, shifted to 22 °C, and induced by 1 mM isopropyl-1-thio-β-D-galactopyranoside overnight at 22 °C. Selenium-containing Δ157 Dot1 protein (with seven methionines) was expressed in a methionine auxotrophic strain of Escherichia coli for 16 h. Preparation of chicken erythrocyte nuclei was performed as described in Ref. 26, except that 0.5% Nonidet P40 was used instead of Nonidet L.E.

Soluble chromatin was isolated from 1 ml of isolated nuclei by treatment with micrococcal nuclease, according to Ref. 27. In brief, isolated nuclei were treated with 20 units of micrococcal nuclease at 37 °C for 30 min in 15 mM Hepes (pH 7.5), 65 mM NaCl, 65 mM KCl, 2 mM MgCl2, 0.2 mM CaCl2, and 2× complete EDTA-free protease inhibitor solution (Roche Diagnostics). Following incubation, EDTA was added to 5 mM, and the reaction cooled on ice. Nuclei were pelleted by centrifugation (4000 × g, 5 min) and broken by Dounce homogenization. After a second centrifugation, chromatin was solubilized in 2 ml of 20 mM Tris (pH 7.5), 600 mM NaCl, 0.5 mM β-mercaptoethanol, and 0.2 mM EDTA. Under these conditions, linker histones dissociate. Mono- and polynucleosomes were isolated by gel filtration in the solubilization buffer over a Sephacryl S-300 column (Amersham Biosciences). DNA content was evaluated by phenol-chloroform extraction and agarose gel electrophoresis. Fractions judged to contain polynucleosomes had a minimum of 800 base pairs of DNA associated with them, whereas mononucleosomes had equal to or less than 200. Nucleosome preparations appeared to be linker-histone and contaminant-free by SDS-PAGE. Fractions were pooled, concentrated against solid sucrose, and dialyzed against 10 mM Tris (pH 7.5), 0.1 mM EDTA, and 0.5 mM β-mercaptoethanol. Nucleosomes were snap frozen and stored at –80 °C.

Crystallography—Purified Δ157 protein was concentrated to 15–20 mg/ml in the gel filtration column buffer with 600 μM AdoHcy. Crystals were obtained via the hanging drop method, with the mother liquor containing 100 mM Hepes (pH 7.0), 18% PEG 8000 (or 20% PEG 2000 MME), and 15% glycerol at 16 °C. The diffraction data were collected at beamline X26C of the National Synchrotron Light Source, Brookhaven National Laboratory. A selenium-containing crystal was used to collect a single wavelength data set at the selenium absorption peak wavelength (Table I). The data were reduced and scaled using DENZO/SCALEPACK package (28). There are three molecules per asymmetric unit. SOLVE (29) readily determined 21 selenium sites (7 per molecule), whose positions were related by 3-fold non-crystallographic symmetry. The phases were greatly improved by RESOLVE (30) and the resulting electron density map was suitable for the initial chain tracing using O (31). Non-crystallographic restraints were imposed on the three protein molecules throughout the refinement by CNS programs (32) against the native data set and were only released from protein side chains at the last cycle to account for different crystal packing environments. After several cycles of refinement and manual rebuilding, the final model was refined with an R value of 21.3% and a free R value of 24.6%. Among the nonglycine and nonproline well defined residues, 92% are in the most favored and 8% in the additional allowed regions of a Ramachandran plot (33). The figures were drawn with the programs XtalView (34), MOLSCRIPT (35), and Raster3D (36). The atomic coordinates have been deposited in the Protein Data Bank with accession code 1U2Z.

RESULTS

A Conserved Dot1p Core—Purified full-length yeast Dot1p (582 residues) was subjected to limited protease digestion by V8 protease, generating a stable C-terminal fragment representing amino acids between 158 and 582, termed Δ157. Dot1p Δ157 contains a core region conserved among human, Caenorhabditis elegans, Drosophila, and Anopheles gambiae Dot1p homologues. The length of these Dot1 proteins varies from 582 amino acids in yeast to 2237 amino acids in Drosophila. The conserved Dot1p core is located at the C terminus in yeast, but is at the N terminus in human, C. elegans, Drosophila, and Anopheles gambiae Dot1p homologues (Fig. 1A). The Δ157 protein, expressed more efficiently than the full-length protein, retained partial activity on nucleosomes (Fig. 2A). Here we describe the structure of Δ157 in complex with the methyl donor by-product AdoHcy.

Overall Structure of Dot1p—Electron density maps were calculated using single wavelength anomalous diffraction data from a SeMet-incorporated crystal (Table I). The crystallographic asymmetric unit contains three molecules of Δ157 (Fig. 3A). However, analytic gel filtration and dynamic light scattering measurements suggest that the protein Δ157 has an apparent molecular weight of a monomer (data not shown) and we thus describe the monomeric structure.
FIG. 1. Conserved Dot1-core region. A, schematic representation of Dot1 homologues from yeast and human. A proteolytic cleavage site of the yeast Dot1 is labeled V8. B, sequence alignment of the conserved Dot1 core region. The alignment includes members of the Dot1 family from various species: S. cerevisiae (NP_010728), C. elegans (NM_085569), Drosophila (AE003875), A. gambiae (EAA05558), mouse (BB678539), and human (NP_115871). The residue numbers and secondary structural elements of the Dot1 core (helices A-M and strands 1–11) are shown above the aligned sequences. Dashed lines indicate disordered regions. The amino acids highlighted are invariant (white letter against black background) and conserved (black against gray) among six members of the Dot1 family. The letters above the sequences indicate the structural/functional roles of the corresponding yeast Dot1 residues: h indicates intramolecular hydrophobic interaction, i indicates intramolecular nonhydrophobic (polar or charge) interaction, s indicates surface-exposed invariant residues potentially important for substrate binding, t indicates structural residue Gly involved in the sharp turn, asterisk indicates cofactor-interacting residue, K indicates residues forming the target Lys binding pocket, and R indicates residues potentially involved in binding an Arg (most likely from the substrate). The conserved sequence motifs characteristic of class I MTases are labeled with roman numbers according to Ref. 39.
Monomeric Δ157 contains two domains (Fig. 3B). The N-terminal domain consists of 4 β-strands (β1–β4) and 6 α-helices (αA–αF). Three helices (αB, αC, and αD), together with a hairpin (β2 and β3), mimic a classic up and down four-helix bundle, where the hairpin replaces the fourth helix. The N-terminal residues (purple), unique to yeast Dot1p, appear to be critical to
the structural integrity of the molecule: the 41-residue segment (amino acids 176–216) forms a V-shaped structure clamped onto nearly the entire N-terminal domain, pairing strand J with helix A, and packing helix A with an open cleft located between the two domains (Fig. 3D). In comparison with human Dot1L (25), a longer and bent helix is the observed intensity and $I$ is the averaged intensity from multiple observations.

$$R_{merge} = \frac{\sum |I - \langle I \rangle|^2}{\sum |I|^2}$$

$$R_{merge}^a = \frac{\sum |F_o - F_c|^2}{\sum |F_o|^2}$$

$$R_{merge}^a$$ was calculated using a subset (5%) of the reflections not used in refinement.

The concave cleft between the two domains is the likely binding site for the substrate. The bottom of the cleft is highly acidic (Fig. 4A), complementing with the basic nucleosomal disk surface around Lys79 of histone H3 (Fig. 4E), which includes many positively charged residues of H3 (Arg32, Arg38, H4 (Arg79, Lys89, and Arg95), and H2B (Arg38, Arg38, and Lys405), forming a core nucleosome surface crucial for transcriptional silencing (38). The Arg- and Lys-binding sites identified on the surface of the concave cleft may provide a docking site for the outwardly protruding basic side chains around Lys79 of histone H3. Many of the residues that form the Arg- and Lys-binding sites, particularly Asp301, Tyr350, Tyr372, and Glu374, are invariant in the six Dot1p homologous shown in Fig. 1B. The importance of these residues in histone binding is supported by site-directed mutagenesis experiments. Changes at the two negatively charged residues (D301A, D301N, E374A, and E374Q) essentially abolished HKMT activity (Fig. 4F, top two panels). However, the mutants had normal expression levels and retained the ability to bind AdoMet, DNA, and nucleosomes (Fig. 4F, three bottom panels). We suggest that the loss of methyl transfer activity was because of defects in binding of one of the basic residues near Lys79 of histone H3 on the disk-like nucleosome surface. Such a defect may affect proper positioning of the target Lys79 into the active site, without total disruption of the nucleosome association through DNA interaction. Of course, we cannot completely exclude the involvement Glu374, located near the active site, in catalysis, although we consider it unlikely (see “Discussion”).

Dot1p Core Contains Conserved Sequence Motifs Common to Class I MTases—Having determined the structure of yeast Dot1p, we were able to perform a structure-guided sequence alignment of Dot1 family proteins (Fig. 1B). The sequence alignment reveals invariant or conservatively substituted positions scattered throughout the conserved core. Only two regions of the conserved core have more than 2-residue insertions or deletions among the different Dot1 homologues (Fig. 1B). In comparison with human Dot1L (25), a longer and bent helix αC is formed (because of Pro439) by yeast Dot1p residues 284–300 and additional secondary structures helix αE and strand β4 are formed by yeast Dot1p residues 325–338.

The four most highly conserved segments involve consecutive invariant residues, namely YGE prior to helix αH, DLGS-GVG in the carboxyl end of strand β5 and the loop that follows, NNF after strand β6, and VSWT between strands β10 and β11. These segments correspond to the sequence motifs X, I, IV, and VIII in the class I MTases; we therefore retain the nomenclature (39). However, the structure shows that these four motifs are clustered together on one surface patch at or near the AdoHcy-binding site (Fig. 3E). In addition, motifs II and III contain a single invariant residue (Glu422 and Phe469) involved in direct interactions with AdoHcy (see below).

The structure suggests that the conserved Dot1p motifs have functional importance. Besides being involved in direct interactions with AdoHcy (motif I, Asp301 and Gly389; motif II, Glu374; and motif III, Phe469), active site formation (motif X, Glu374, motif IV, Phe481; and motif VIII, Trp512), and catalysis of methyl transfer (motif IV, Asn527), many invariant residues
Fig. 3. Structure of yeast Dot1p. A, a trimer of Δ157 formed in the crystallographic asymmetric unit. The trimer interface is formed between the N-terminal helix αA (which is unique to yeast Dot1p and C. elegans homologue, see Fig. 1B) and the outer surface of the AdoHcy-binding site of a neighboring non-crystallographic related molecule. We also note that yeast Hst2, a member of Sir2 family of NAD^+-dependent protein deacetylases, forms a functional homotrimer mediated by the N-terminal extension to the central catalytic core domain (62), which is characteristic of the other Sir2 homologues. On the other hand, helix αA may be involved in regulation of Dot1p via protein-protein interactions. B, monomer structure of Δ157. The N-terminal clamp is shown in purple, the N-terminal helical domain in green, and the C-terminal catalytic domain in orange. The bound AdoHcy is shown as a stick model. C, the polar interface between the N- and C-terminal domains. D, a GRASP (63) representation showing the cleft between the two domains and a surface hole, through which the adenine ring of AdoHcy is visible. E, a GRASP representation, 90° rotated from the view in panel D, showing the conserved motifs (X, IV, and VIII) surrounding the active site pocket, through which the AdoHcy sulfur atom is visible. Conserved motifs I, II, and III are buried and invisible from the surface. In addition, there are three surface-exposed invariant residues, Tyr364 and Phe367 in the loop after helix αG, and Lys508 after strand s9 (Fig. 1B). The surface locations of Tyr364 and Phe367, approximate to the motif X that provides residues in the active site formation (Val371 and Gly373, Fig. 4B), the Arg-binding site (Tyr372 and Glu374, Fig. 4C), and the lid to the AdoHcy binding site (Fig. 5, C and E), suggests these two conserved residues may have implications in specific enzyme-substrate interactions. Lys508, located right next to Phe367 of motif IV, might also contribute to substrate binding and/or catalysis. F, a model of Dot1p docked with a nucleosome. The structure of the nucleosome core particle is shown as ribbons (red, H3; green, H4; magenta, H2A; yellow, H2B; gray lines, DNA). The model was put together by aligning the target Lys79 of histone H3, located on the nucleosome disk surface, with the active site pocket of Dot1p. By rotating the Dot1p along the Lys79 active site alignment, the N-terminal residues, either disordered (residues 158–175) or not included in the crystallization (residues 1–157), could be in a position to contact DNA directly and/or involved in interaction with monoubiquitin covalently attached to histone H2B Lys123. In the model shown, no conformational changes are required to dock. However, major conformational changes of either the nucleosome and/or the enzyme would be needed to allow the side chain of Lys79 of H3 moving additional a few angstroms into the active site pocket.
(motif I, Ser\textsuperscript{400}; motif IV, Asn\textsuperscript{460}, and motif VIII, Ser\textsuperscript{542}) are involved in intramolecular interactions that likely confer stability to the molecule, particularly around the AdoHcy-binding and active sites.

**Dot1p-AdoHcy Interactions**—In the binary Δ157-AdoHcy complex, the methyl-donor product AdoHcy (added during purification, see “Experimental Procedures”) is observed at the carboxyl end of the parallel strands β₅, β₆, β₇, and β₈ (Fig. 3B), the hallmark of a nucleotide-binding site. The AdoHcy in yeast Dot1p (as well as the AdoMet in human Dot1L) is in an extended conformation (Fig. 5A), most frequently observed in widespread class I MTases such as the DNA cytosine MTase DNMT2 (40), and protein-arginine MTase PRMT1 (41). However, the AdoHcy/AdoMet conformation in the Dot1 proteins is significantly different from the folded conformation observed in the SET domain of HKMTs (19, 20, 42–45). Such different conformations of the cofactor may provide a good target to design inhibitors that are selective for class I (Dot1p and PRMT1) versus class V (SET HKMTs) MTases.

Dot1p-AdoHcy interactions can be grouped according to the three moieties of AdoHcy (Fig. 5B). Invariant residues from motifs I, II, and III provide most of the specific interactions; these interactions are buried and invisible from the surface (Fig. 3E). (i) The motif I Gly-rich loop (GSGVG) between strand β₅ and helix α₇ bends sharply underneath the homocysteine moiety. The last residue of strand β₅, Asp\textsuperscript{387} of motif I, interacts with the amino group (NH\textsubscript{2}) of the AdoHcy. In addition, Gly\textsuperscript{389} (motif I) makes a van der Waals contact with the ribose ring oxygen O₄. (ii) A strongly conserved acidic residue (Glu\textsuperscript{422}) at the carboxyl end of strand β₆ (motif II) forms bifurcated hydrogen bonds with the ribose hydroxyls; this interaction is almost universal in class I MTases (14). An E422A mutation abolishes both AdoMet binding (measured by cross-linking) and MTase activity, whereas an E422D mutation retains full activity (Fig. 2G). (iii) The adenine ring is flanked via van der Waals contacts by the phenyl ring of Phe\textsuperscript{460} after strand β₆ (motif III) and Ile\textsuperscript{463} after strand β₆ (motif II). The exocyclic amino group (N₆) of adenine makes a hydrogen bond with Ser\textsuperscript{465} (motif III).

A unique interaction of Dot1p involves motif X, in which two Tyr residues (370 and 372) are involved in water-mediated interactions with AdoHcy. The hydroxyl oxygen of Tyr\textsuperscript{370} shares a water molecule with N₆ and the ring nitrogen N₁ of adenine, whereas the hydroxyl oxygen of Tyr\textsuperscript{372} shares a water molecule with one of the AdoHcy carboxyl oxygen atoms (COO⁻) (Fig. 5B). The two Tyr aromatic rings effectively provide a lid for the entrance of cofactor binding site and nearly bury AdoHcy (Fig. 3, D and E). On the surface of the molecule, there are two holes through which AdoHcy is visible: one is in the active site where the AdoHcy sulfur atom is visible (Fig. 3E), the other is on the back of the molecule where the AdoHcy adenine ring is visible (Fig. 3D). Both holes are too small to let AdoHcy diffuse out from the binding site, which may explain the observation that AdoMet is co-purified with human Dot1L (25). The buried AdoHcy suggests that exchange between the methyl donor AdoMet and the reaction by-product AdoHcy requires the movement of the lid, formed by the highly conserved motif X residues and the associated region that are part of the active site (Val\textsuperscript{371} and Gly\textsuperscript{373}, Fig. 4B) as well as the potential substrate Arg-binding site (Tyr\textsuperscript{72} and Glu\textsuperscript{74}, Fig. 4C). The fact that a mixture of unmodified, mono-, di-, and trimethylated H3 Lys\textsuperscript{79} coexists in yeast (8) suggests that the exchange of the reaction product AdoHcy with AdoMet in the closed-lid binding site would require the release of the substrate and therefore should require methyl transfer to proceed distributively.

**The Active Site and Enzymatic Properties of Yeast Dot1p**—The active site is situated in a very narrow surface pocket (Fig. 4B), where the AdoHcy sulfur atom lies at the bottom and is visible (Fig. 5C). The opening of the pocket is ~4 × 5 Å, a dimension that would barely accommodate the side chain of a lysine such that the terminal amino group could reach the sulfur atom of AdoHcy, where the transferable methyl group would be attached in AdoMet. The residues forming the sides of active site pocket are Val\textsuperscript{371} and Gly\textsuperscript{373} of motif X, Phe\textsuperscript{461} and Leu\textsuperscript{462} of motif IV, and Trp\textsuperscript{463} of motif VIII (Fig. 4B). The hydrophobic nature of these residues corresponds with the four hydrophobic methylene groups of a lysine side chain. Replacement of Trp\textsuperscript{463} to Phe or Ala nearly abolishes MTase activity, but not binding of AdoMet (measured by cross-linking), nucleosomes, or DNA, indicating that the indole ring is most likely involved in target lysine binding (Fig. 2G).

In the absence of the target lysine, four inter-connecting water molecules (w1–w4) fill the active site pocket, three of which are caged by backbone carbonyl oxygen atoms of Asn\textsuperscript{469}, Val\textsuperscript{371}, and Asn\textsuperscript{479} (Fig. 5C). The fourth water molecule (w4) is held in place between the face of the indole ring of Trp\textsuperscript{463}, the edge of the phenyl ring of Phe\textsuperscript{461}, and the aliphatic Leu\textsuperscript{462}. The side chain of Asn\textsuperscript{379} is the only side chain that protrudes into the active site from the bottom. Mutations of the equivalent residue in human Dot1L, Asn\textsuperscript{441}, to Asp or Ala abolished activity (25). An Asn side chain of the so-called NPPY motif is used in the active site of amino MTases of adenine or cytosine in DNA and of protein glutamine MTases (46–48). Remarkably, using AdoHcy as an anchor point, the invariant Asn\textsuperscript{479} of Dot1p is superimposable onto the corresponding Asn in TaqI DNA adenine MTase (not shown) and HemK protein glutamine MTase (Fig. 5D). The target amino nitrogen atom occupies the position of the water molecule (w3), which hydrogen bonds to the side chain of Asn\textsuperscript{479} and the backbone carbonyl oxygen of Val\textsuperscript{371}. This suggests a potential similarity in the catalytic mechanism between Dot1p and class I amino MTases. In the latter case, the amino group (NH₃⁺) that becomes methylated is not charged and is positioned for an in-line attack on AdoMet. However, the amino group (NH₃⁺) of a Lys side chain is usually positively charged. Under laboratory conditions, Dot1p is active in a broad pH range, from pH 6 to 9.5 (and beyond) with a maximum activity around pH 8.5, coinciding with strongest cross-linking to AdoMet (Fig. 2F). This is very different from SET domain containing HKMTs such as DIM-5 and SET7, which has a narrower pH range (active at pH 8 or higher) and an unusually high pH optimum (~10) (42, 49). At pH 10, the amino group of the target lysine should be partially deprotonated. Only the deprotonated target Lys has a free lone pair of electrons capable of nucleophilic attack on the AdoMet methyl group. Dot1p must use a different mechanism to deprotonate the target lysine (see “Discussion”).

Interestingly, the largest structural difference between yeast Dot1p and human Dot1L in the core region lies in the loop between strands β₁₀ and β₁₁. This loop adopts completely different conformations in the two structures (Fig. 5E), resulting in a closed active site pocket in yeast Dot1p and an open one in human Dot1L (Fig. 5F). In yeast Dot1p, the active site pocket in each of three monomers is identical, even though each monomer has a different crystal packing environment. The loop containing motif VIII is highly conserved both in length and amino acid identity, including four consecutive invariant residues: VSWT (see Fig. 1B). Whereas the corresponding loop in human Dot1L is not well ordered, having high temperature factors (25), the loop in the yeast enzyme is highly structured with many stabilizing intramolecular interactions. For example, both the side chain (O-γ) and main chain (N-H) of Ser\textsuperscript{542} of
motif VIII interact with main chain atoms (N-H and C=O) of Glu^{374} of motif X (Fig. 5F). These interactions link two highly conserved sequence motifs (X and VIII) together to form the closed active site. As mentioned earlier, Glu^{374} interacts with an Arg (Fig. 4C). We suggest that the binding of this Arg (whether it is from the neighboring molecule in the crystal or
Fig. 5. **AdoHcy binding site and active site.**

**A,** superimposition of AdoHcy conformations in yeast Dot1p and SET domain HKMT DIM-5 (19). **B,** stereo view of the AdoHcy binding site. The difference electron density map is contoured at 2.5σ. **Dashed lines** indicate hydrogen bonds. The letter *w* indicates bound water molecules. Interactions involving main chain (mc) atoms are indicated in parentheses. **C,** stereo view of active site. The AdoHcy is in van der Waals surface model, colored with red for oxygen atoms, blue for nitrogen atoms, white for carbon atoms, and green for the sulfur (placed where the transferable methyl group would be attached in AdoMet). **D,** superimposition of yeast Dot1p (AdoHcy and Asn479) and protein glutamine MTase HemK (48) (AdoMet, Asn197, and free glutamine). **E,** structural comparison of yeast Dot1p (residues 176–567; top panel) and human Dot1L (residues 5–332; middle panel). Besides deletion or insertion (located in N and C termini), the largest conformational differences, indicated by arrows, between the two catalytic domains, are the hairpin loop between strands β10 and β11 containing motif VIII and the lid for the entrance to the cofactor binding (containing motif X). **F,** superimposition of yeast Dot1p and human Dot1L (bottom panel), indicating a hinge movement between two domains, possibly caused by the binding of the Arg (see Fig. 4C). **G,** superimposition of the hairpin loop conformations in yeast Dot1p and human Dot1L (in parentheses).
from the nucleosome-disk surface in solution) promotes the interactions between motifs X and VIII. These interactions result in the hinge movement between the two domains (Fig. 5E) and bring in Trp1543, an important residue for the formation of the ordered active site. Substrate induced conformational changes have been observed in many enzymes including the DNA MTase (50) (hinge movement) and in the SET domain HKMT DIM-5 (19) (formation of an ordered active site). The conformational change of the loop containing highly conserved motif VIII suggests that this part of Dot1p is important for substrate binding and/or regulation of its activity by adopting different conformations. Tyr550 of strand β11 adopts a different rotamer conformation from the equivalent Tyr512 in human Dot1L (Fig. 2F). Conservative change of Y550F, or Y312F in human Dot1L (25), retained normal activity, whereas Y550A, or Y312A in human Dot1L (25), abolished activity (Fig. 2G), indicating the importance of maintaining the hydrophobic core between strand β11 and helix αH.

The Segment Involved in Nucleosomal DNA Interaction Is Disordered—The first 15 residues of Δ157 (158–172), disordered in the current structure, contain 6 Lys and 2 Arg. We generated a construct, Δ172, that lacks this positively charged region, and found that it is completely inactive on nucleosomes, whereas full-length Dot1p and Δ157 are both active (Fig. 2A). Δ172 retains the ability to bind AdoMet as measured by cross-linking (Fig. 2B), but loses the ability to bind nucleosomes or 36-bp duplex DNA (Fig. 2C). This observation suggests that this disordered, positively charged region is involved in contacting nucleosomal DNA, and essential for Dot1p activity. Similarly, a stretch of positively charged residues at the C terminus of human Dot1L was demonstrated to be critical for nucleosome binding and therefore, enzymatic activity (25).

In agreement with published data (8–11), Dot1p is a nucleosome-dependent HKMT, i.e. Dot1p is inactive on histones alone (Fig. 2, A and D). However, preincubation of DNA, either the 150- (Fig. 2D) or 30-base pair duplex (data not shown), with Dot1p stimulates its HKMT activity on histones to almost nucleosomal levels. As expected, histone H3 is the target of methylation for both the nucleosomal substrate and the DNA/histone mixture (Fig. 2D). In all cases, Δ157 has reduced activity compared with that of full-length Dot1p, probably because Δ157 lacks additional N-terminal positively charged residues between amino acids 106 and 157 (Fig. 1A).

DISCUSSION

Deprotonation of Target Lys—Dot1p must provide a local environment for the target lysine that will enable it to remain deprotonated. There are many examples of enzymes that contain lysine residues with significantly depressed pKₐ values. A lysine with a low pKₐ is generated when a positive charge is immediately proximal to the lysine (51). The proximity of methylation target Lys to the positively charged methylsulfonium group of AdoMet could have a similar effect. A hydrophobic microenvironment is another situation that produces a lowered pKₐ value. There are examples of buried lysines with pKₐ values as low as 6.5 (53). This scenario can best explain the activity of Dot1p over a wide pH range. The active site pocket sided with all hydrophobic residues, in conjunction with a positively charged methylsulfonium group sitting at the bottom of the pocket, is probably essential for lowering the pKₐ of the ε-amino group of the target Lys so that it can stay in the deprotonated state required for its methylation.

A glutamate residue was suggested to be a general base for catalysis of histone lysine acetyltransferases (54, 55). Specifically, the Glu is located at the bottom of the substrate-binding cleft and surrounded by several non-polar residues that could raise the pKₐ of the glutamate side chain and thus facilitate its ability to extract a proton. A water molecule is bound between the carboxylate and the target Lys to shuttle a proton. There is no equivalent carboxylate group in the structure of yeast Dot1p presented here. The side chain of Glu123⁴, the only negatively charged residue near the active site, is located on the surface of the protein and is kept away from the active site (by two polar groups) to maintain the hydrophobic nature of the microenvironment. Another reason to keep Glu123⁴ away from active site is to avoid an increase in a lysine side chain pKₐ that can occur when there is a carboxylate nearby (56). The pKₐ of the ε-amino group increases, as a result of its greater affinity for protons, to neutralize the negatively charged carboxylate. A perceived conformational change that flips the side chain of Glu123⁴ toward the active site would result in the interaction with the target Lys and/or positively charged methylsulfonium group of AdoMet, both interactions would slow the deprotonation of the target Lys or methyl transfer from the AdoMet.

Does Dot1p Require Ubiquitination of Histone H2B Lys79 in Vivo?—In S. cerevisiae, methylation of Lys79 of H3 requires ubiquitination of Lys123 of histone H2B (11, 22). H2B Lys123 is located on the same nucleosome disk surface (Fig. 3F), ~30 Å away from the target Lys79 of histone H3. Contrary to the in vitro data, recombinant Dot1p (both the full-length and Δ157 proteins) was active on nucleosomes assembled in vitro from bacterially expressed, recombinant core histones (a gift of Dr. K. Lugner) (Fig. 2A, right panel), indicating that ubiquitination is not required for Dot1p activity in vitro. Interestingly, a stretch of ~60 amino acids (residues 40–100) of yeast Dot1p, not included in the crystallization, has repeated hydrophobic residues every 4–5 positions, and is predicted to form short helices by secondary structure prediction (57). This stretch of Dot1p is similar in size to the ~50 amino acid monoubiquitin-binding domain (CUE) and the ubiquitin-binding UBA domain, both of which have a three-helix bundle structure (58–61). It is possible that Dot1p interacts, via this region, directly with H2B ubiquitinated nucleosome or indirectly through other ubiquitin-binding proteins. Such an interaction could be significant in vivo, recruiting Dot1p to specific high-order chromatin where ubiquitinated histone H2B might serve as a spacer between adjacent nucleosome disk surfaces, allowing Dot1p access to its target Lys (24).

Conclusions—We have described the crystallographic structure of the conserved region of yeast Dot1p. This region of Dot1p is responsible for cofactor binding and catalysis of methyl transfer. The elongated structure has a large acidic cleft that is the likely binding site for the basic nucleosome disk surface centered around Lys79 of histone H3. Although the majority of the conserved Dot1p core is highly structured, a positively charged N-terminal region, disordered in the current structure, is important for catalysis, apparently via the interaction with the nucleosomal DNA that is wrapped around histones. We suggest 1) the target amino group of Lys79 of H3 is positioned by the conserved Asn79 of motif IV such that the lone pairs of the nucleophilic nitrogen point toward the incoming methyl group and the charged sulfonium ion of AdoMet, based on the location of Asn79, structural comparison to other class I MTases, and the pH requirements; 2) Dot1p is a catalytic enzyme, based on the fact that a mixture of unmodified, mono-, di-, and trimethylated H3 Lys79 coexists in yeast and the exchange of the reaction by-product AdoHcy with AdoMet in the closed-lid binding site; 3) the highly conserved motif VIII is important for substrate binding and/or regulation of Dot1p activity by adopting a different hairpin-loop conformation between strands β10 and β11; and 4) the N-terminal helix αA, unique to yeast Dot1p and C. elegans homologue, may be involved in regulation of Dot1p via protein-protein interactions.
Further structural and biochemical studies of the yeast Dot1p-nucleosome complex are needed for understanding the nature of Dot1p-nucleosome interactions and the molecular mechanisms of nucleosomal histone methylation and its dependence on ubiquitin in vivo.

Acknowledgments—We thank Drs. Fred van Leeuman and Daniel E. Gottschling for the full-length yeast Dot1, Dr. Karolin Luger for the recombinant nucleosomes, and Dr. Venki Ramakrishnan for help with x-ray data collection at beamline X26C in the facility of the National Synchrotron Light Source, and Dr. Fred van Leeuman for insightful comments on the manuscript.

REFERENCES

1. Zhang, Y., and Reinberg, D. (2001) Genes Dev. 15, 2343–2360
2. Lachner, M., O’Sullivan, R. J., and Jenuwein, T. (2003) J. Cell Sci. 116, 2117–2124
3. Shiio, Y., and Eisenman, R. N. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 13225–13230
4. Strahl, B. D., and Allis, C. D. (2000) Nature 403, 41–45
5. Jenuwein, T., and Allis, C. D. (2001) Science 293, 1074–1080
6. Fischle, W., Wang, Y., and Allis, C. D. (2003) Nature 423, 475–479
7. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khosravanizadeh, S. (2003) Genes Dev. 17, 1870–1881
8. van Lueeune, P., Gaaren, P. R., and Gottschling, D. E. (2002) Cell 109, 745–756
9. Lacoste, N., Ulej, R. T., Hunter, J. M., Poirier, G. G., and Cote, J. (2002) J. Biol. Chem. 277, 30421–30424
10. Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002) Curr. Biol. 12, 1052–1058
11. Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002) Genes Dev. 16, 1518–1527
12. Singer, M. S., Kahn, A. J., Meisinger, L. L., Peterson, S. E., Goggin, C., Mahowald, M., and Gottschling, D. E. (1988) Genetics 116, 613–632
13. Dlakic, M. (2001) Trends Biochem. Sci. 26, 405–407
14. Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Trends Biochem. Sci. 28, 329–335
15. Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., and Cheng, X. (2003) Curr. Biol. 13, 1298–1303
16. Fang, J., Feng, Q., Ketel, C. S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., and Allis, C. D. (2002) Mol. Cell Biol. 22, 1288–1306
17. Fauch, J., Feng, Q., Ketel, C. S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J. A., and Zhang, Y. (2002) Curr. Biol. 12, 1086–1099
18. Nishioka, K., Rice, J. C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chui, S., Valenza, P., Tempst, P., Steward, R., Lis, J. T., Allis, C. D., and Reinberg, D. (2002) Mol. Cell 9, 1201–1213
19. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Meller, J., and Kozarides, T. (2002) Nature 419, 407–411
20. Zhang, Y., Yang, Z., Khan, S. I., Horton, J. R., Tamura, H., Selker, E. U., and Cheng, X. (2003) Mol. Cell 12, 177–185
21. Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, J. A., Blackburn, G. M., and Gamblin, S. J. (2003) Nature 421, 652–656
22. Tamara, H., Zhang, X., Millen, D., Singh, P. B., Nakayama, J., Grewal, S. I., Allis, C. D., Cheng, X., and Selker, E. U. (2003) Nat. Genet. 34, 75–79
23. Briggs, S. D., Xiao, T., Sun, Z. W., Caldwell, J. A., Shabeanworite, J., Hunt, D. F., Allis, C. D., and Struhl, B. D. (2002) Nature 418, 498
24. Dover, J., Schneider, J., Tawiah-Baaheng, M., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002) J. Biol. Chem. 277, 28368–28371
25. Sun, Z. W., and Allis, C. D. (2002) Nature 418, 104–108
26. Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R. M. (2003) Cell 112, 711–725
27. Kong, R. S., Daniels, C. M., Francis, S. A., Shih, S. C., Salerno, W. J., Hicke, L., and Feigon, J. (2003) Cell 112, 621–630
28. Mueller, T. D., and Feigon, J. (1998) J. Mol. Biol. 283, 1042–1047
29. Zhao, K., Chai, X., Clements, A., and Marmorstein, R. (2003) Nat. Struct. Biol. 10, 864–871
30. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296