(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2012 (26.04.2012)

PCT

(10) International Publication Number
WO 2012/052391 A1

(51) International Patent Classification:
G06F 19/16 (2011.01)

(21) International Application Number:
PCT/EP2011/068087

(22) International Filing Date:
17 October 2011 (17.10.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/394,543 19 October 2010 (19.10.2010) US

(71) Applicant *(for all designated States except US)*: **GLAXO GROUP LIMITED** [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford Middlesex UB6 0NN (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHUNG, Chun-Wa** [GB/GB]; GlaxoSmithKline, Gunnels Wood Road, Stevenage Hertfordshire SG1 2NY (GB). **MOSLEY, Julie** [GB/GB]; GlaxoSmithKline, Gunnels Wood Road, Stevenage Hertfordshire SG1 2NY (GB). **THOMAS, Pamela Joan** [GB/GB]; GlaxoSmithKline, Gunnels Wood Road, Stevenage Hertfordshire SG1 2NY (GB).

(74) Agent: **LEAROYD, Stephanie Anne**; GlaxoSmithKline, Global Patents (CN925.1), 980 Great West Road, Brentford Middlesex TW8 9GS (GB).

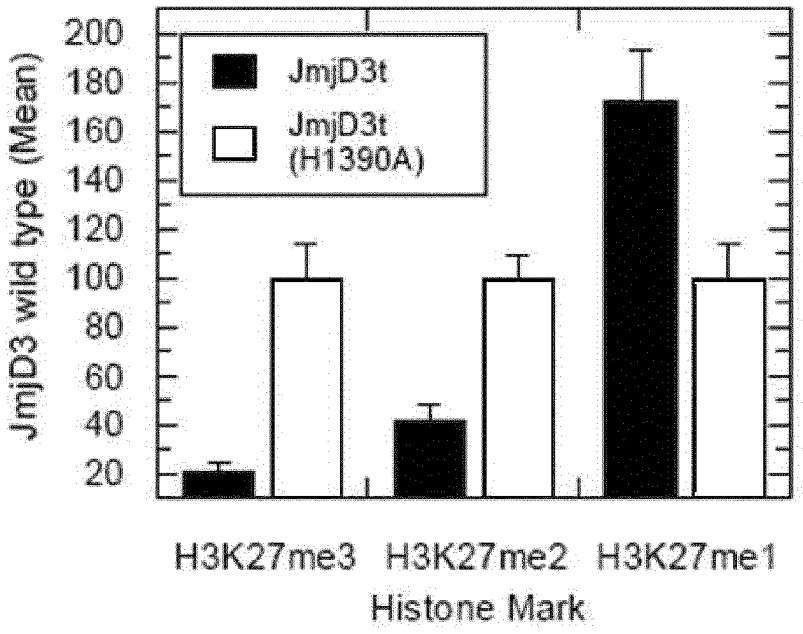
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ,

[Continued on next page]

(54) Title: POLYPEPTIDE WITH JMJD3 CATALYTIC ACTIVITY

FIG. 6



(57) Abstract: Polypeptides containing JMJD3 catalytic activity, amino acid and nucleotide sequences encoding the same, and uses and the crystal structure thereof, are provided.

WO 2012/052391 A1

DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

Polypeptide with JMJD3 Catalytic Activity

Field of the Invention

The present disclosure provides, *inter alia*, polypeptides that comprise JMJD3 (jumonji domain containing 3) catalytic activity, nucleic acid sequences encoding the same, and crystal structure and uses thereof. Also provided herein are polypeptides that comprise a novel GATA-like motif, sequences encoding the motif, and uses thereof.

Background of the Invention

The structure of chromatin is complex and dynamic and has a major effect on gene transcription. Chromatin is the complex combination of DNA and protein that makes up chromosomes. It is found inside the nuclei of eukaryotic cells and is divided between heterochromatin (condensed) and euchromatin (extended) forms. The major components of chromatin are DNA and proteins, including histones. The basic building blocks of chromatin are nucleosomes, each of which is composed of 146 base pairs of DNA wrapped around a histone octamer consisting of 2 copies of each H2A, H2B, H3 and H4. The functions of chromatin are to package DNA into a smaller volume to fit in the cell, to strengthen the DNA to allow mitosis and meiosis, and to serve as a mechanism to control expression and DNA replication. Chromatin contains genetic material serving as instructions to direct cell functions. Changes in chromatin structure are regulated by modifications on histone and DNA methylation. Epigenetic mechanisms do not change the DNA sequence but allow the genes to be expressed differently. Epigenetic modifications include numerous mechanisms including DNA methylation and post-translational modification of N-terminal tails of histone proteins such as methylation, acetylation, phosphorylation and ubiquitination.

In recent years many of the target proteins involved in epigenetic control have been identified. These targets can be classified into families that read, write, and erase covalent modification (e.g., methylation, acetylation) in histones.

Histone methylation is an abundant epigenetic modification of core histones found in eukaryotic organisms that has been linked to a number of cellular processes including DNA repair, cell cycle progression, cell differentiation and regulation of gene expression. This modification is catalysed by the specific histone methyltransferases (HMTs), lysine methyl transferase and arginine methyl transferase, which introduce methyl groups at lysine (K) or arginine (R) residues respectively. Thus far, histone methylation has been found to occur at six major sites, including histone H3 lysine 4 (H3K4), H3K9, H3K27, H3K36, H3K79 and H4K20. Unlike other modifications, the same lysine residue can be methylated to different degrees to include mono-, di- or trimethyl moieties, which may have different functional consequences. In general, lysine methylation at H3K4, H3K36 and H3K39 is associated with regions of transcriptionally active chromatin, whereas methylation at H3K9, H3K27

and H4K20 is associated with transcriptionally silenced regions (Martin C. and Zhang Y., Nature Rev. Mol. Cell Biol. 2005, 6, 838-849). H3K9 promoter methylation is considered a repressive mark for euchromatic genes (Nielsen et al., Nature 2001,412, 561-565; Shi et al., Nature 2003, 422, 735-738) and is also one of the landmark modifications associated with heterochromatin (Nakayama et al., Science 2001, 292, 110-113) but some studies have also identified association of H3K9 trimethylation (H3K9me3) with actively transcribed genes (Vakoc C. et al, Mol. Cell 2005, 19. 381-391).

Unlike other histone modifications such as acetylation, methylation used to be regarded as a permanent/irreversible modification. However, with the identification of histone demethylases, this process has been shown to be reversible and dynamically controlled. Histone methylation appears thus to be regulated by a complex network that involves a large number of site-specific methylases, demethylases, and methyl recognition proteins, which play an important role in controlling the expression of genetic information through transcriptional changes and chromatin structure alterations. Since levels of lysine methylation are known to change during processes such as transcriptional regulation, it was proposed that specific enzymatic activity might remove the methyl groups (Bannister *et al.*, 2002 Cell 109, 801-806). Recent work has confirmed the existence of enzymatic demethylation and two separate mechanisms of lysine demethylation have been demonstrated: amine oxidation by Lysine Specific Demethylase 1 (LSD1) and hydroxylation by JmjC-domain containing proteins, which indicate these proteins as being novel histone modifying enzymes that can remove methyl groups on lysines (Shi *et al.*, 2004, Cell 119, 941-953; Tsukada *et al.*, 2006, Nature 439, 811-816).

The Jumonji protein is the founding member of a group of proteins characterised by a novel structural motif, the JmjC domain. This is an extensive group of demethylase enzymes which can be defined into several families according to sequence similarity within the JmjC domain and the presence of other domains in the full length protein. The JmjC domain of several members of this family has been shown to possess lysine demethylation activity, which is dependent on iron (Fe (II)) and α-ketoglutarate as co-factors (Klose RJ *et al*, Nat Rev Genet. 2006 Sep; 7(9); 715-27). Unlike LSD1, which can only remove mono- and dimethyl lysine modifications, the JmjC-domain-containing histone demethylases (JHDMs) can remove all three histone lysine-methylation states.

Some JmjC domain-containing proteins, including the histone demethylase JMJD3, have been implicated in tumorogenesis and thus have identified histone demethylases as targets of research for anti-cancer therapies (see for example WO2009/114011 and WO2010/043566).

JMJD3 (KDM6B) is one of the approximately 30 JmjC family members found in humans, and functions as a specific demethylase of lysine 27 of histone H3 (H3K27). JMJD3 can demethylate both the tri- and dimethylated H3K27-repressive histone marks, thereby facilitating gene transcription. This was first demonstrated in C. Elegans embryogenesis, where JMJD3 was shown to regulate gonadal development through modulation of HOX gene expression (Agger K *et al*, Nature 2007 Oct; 449(7163); 731-734). Further studies have placed JMJD3 at key cell fate decision checkpoints in T

lymphocytes (Miller SA *et al*, Genes Dev. 2008 Oct; 22; 2280-2993) and macrophages (Ishii M *et al*, Blood 2009 Oct; 114(15); 3244-3254). In addition, JMJD3 has been demonstrated to regulate the differentiation state of the epidermis (Sen GL *et al*, Genes Dev. 2008 Jul; 22; 1865-1870) and to activate the tumour suppressor, INK4A-Arf, in response stress induced signals (Agger K *et al*, Genes Dev. 2009 Apr; 23; 1171-1176). JMJD3 also appears to be involved in more acute, externally-driven, inflammatory processes. In macrophages, for example, JMJD3 is rapidly induced through an NF-kB-dependent mechanism in response to bacterial products and inflammatory stimuli (De Santa F *et al*, Cell 2007 Sept; 130; 1083-1094). Moreover, depletion experiments in these cells have demonstrated that JMJD3 participates directly in the inflammatory transcriptional response, although it remains unclear whether this is achieved through demethylation of H3K27me3 at target gene promoters (De Santa F *et al*, EMBO J. 2009 Sept; 28; 3341-3352).

Summary of the Invention

Carboxy terminal fragments of the human JMJD3 protein that retain JMJD3 catalytic activity and specificity, and polypeptides containing the same, are described herein. This fragments can be used, *inter alia*, in assays to identify compounds that modulate (e.g., increase or decrease) JMJD3 catalytic activity. This disclosure also relates to (i) crystals of the polypeptides that contain JMJD3 catalytic activity (e.g., human H3K27 demethylase activity) which are suitable for structure determination of the polypeptide using X-ray crystallography and (ii) computational methods using the structural coordinates of the polypeptide to screen for and design compounds that modulate, e.g., decrease (e.g., inhibit) or increase, JMJD3 catalytic activity (e.g., human H3K27 demethylase activity) of the polypeptide. In addition, this disclosure relates to methods for identifying compounds that bind to the polypeptide and/or modulate JMJD3 catalytic activity (e.g., human H3K27 demethylase activity), e.g., in a high throughput setting.

Also provided herein is a novel GATA-like Zn++ binding domain and polypeptides containing the same. In addition, this disclosure provides methods for identifying compounds that bind to the polypeptide and/or modulate Zn++ binding by the polypeptide, e.g., in a high throughput setting.

In one aspect, the disclosure relates to a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. For example, with reference to (iii), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 253-1682 of SEQ ID NO:1; 281-1682 of SEQ ID NO:1; residues 400-1682 of SEQ ID NO:1; residues 600-1682 of SEQ ID NO:1; residues 1000-1682 of SEQ ID NO:1; residues 1100-1682 of SEQ ID NO:1; residues 1140-1682 of SEQ ID NO:1; and so forth. As another example, with reference to (iv), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 247-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 253-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 281-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 400-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 600-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1000-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1100-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1140-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); and so forth. A sequence that is at least 85% identical to a sequence in (i)-(iv) can be, e.g., at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the recited sequence.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is in soluble form.

In some embodiments, the polypeptide is in crystal form. In some embodiments, the crystal has the structure deposited in Protein Data Bank (PDB) ID: 2XUE.

In one embodiment, the polypeptide is suitable for crystallization, i.e., preferably the polypeptide fragment is crystallizable. Preferably, the crystals obtainable from the polypeptide described herein are suitable for structure determination of the polypeptide using X-ray crystallography. Preferably, the crystals are radiation stable enough to permit more than 85% diffraction data completeness at resolution of preferably 3.5 Å or better to be collected upon exposure to monochromatic X-rays.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In another aspect, the disclosure relates to an isolated nucleic acid encoding an isolated polypeptide described herein.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising

culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In one aspect, the disclosure relates to a nucleic acid (e.g., an isolated nucleic acid), wherein the nucleic acid comprises a JMJD3 nucleic acid sequence, wherein the encoded polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 nucleic acid sequence consists of:

(i) a sequence at least 85% identical to nucleotides 3797-5422 of SEQ ID NO:2;

(ii) a sequence at least 85% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2;

(iii) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1115 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. When used for protein expression, the nucleic acid, e.g., with respect to (iii) and (iv), is in the proper reading frame to encode a polypeptide that possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)).

In some embodiments, the JMJD3 nucleic acid sequence consists of (i) a sequence at least 85% identical to nucleotides 3797-5422 of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to nucleotides 3797-5422 of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (ii) a sequence at least 85% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of nucleotides 1133 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1115 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of nucleotides 1115 to 3796 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 nucleic acid sequence).

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 nucleic acid sequence) at 5' or 3' end. For example, the heterologous sequence can include a sequence that encodes a peptide used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a sequence encoding a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag, e.g., at its 3' end. In some embodiments, the nucleic acid comprises a sequence encoding a FLAG tag, e.g., at its 5' end. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag and a FLAG tag, e.g., at its 5' or 3' end. In some embodiments, the nucleic acid comprises a sequence encoding peptides used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the encoded polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the encoded polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for identifying a compound which modulates (e.g., increases or decreases) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:
(a) constructing a computer model of a polypeptide comprising a JMJD3 sequence defined by the structure coordinates of the polypeptide, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

(b) selecting a test compound by a method selected from the group consisting of:

(i) assembling molecular fragments into the compound, (ii) selecting a compound from a small molecule database, and (iii) de novo ligand design of the compound;

(c) employing computational methods to perform a fitting program operation between computer models of the test compound and the polypeptide in order to provide an energy-minimized configuration of the compound with the polypeptide; and
(d) evaluating the results of the fitting operation to quantify the association between the test compound and the polypeptide computer model, thereby evaluating the ability of the test compound to associate with the polypeptide.

In some embodiments, the structure coordinates are deposited in Protein Data Bank (PDB) ID: 2XUE.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some

embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the method further comprises contacting the test compound with the polypeptide or a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide) (e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri-methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri-methylated, at K27)); the substrate can be naturally occurring in the host cell) (e.g., wherein the contacting occurs in conditions that are favorable for JMJD3 catalytic activity in the absence of a test compound); and

evaluating if the test compound modulates JMJD3 catalytic activity of the polypeptide (e.g., as compared to a control). A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

In another aspect, the disclosure relates to a method for identifying compounds which modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:

(a) contacting (e.g., in vitro) a test compound with

(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide)

(e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell); and

(b) analyzing the ability of the test compound to modulate JMJD3 catalytic activity (e.g., H3K27 demethylase activity) of the polypeptide. A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell).

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating (e.g., *in vitro*) a test compound for its ability to modulate (e.g., decrease or increase) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:
(a) incubating a test compound with
(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide),
(e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell)
wherein the incubating occurs in conditions that are favorable for JMJD3 catalytic activity in the absence of a test compound; and
(b) evaluating whether the test compound modulates JMJD3 catalytic activity, e.g., as compared to a control. A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating a test compound, the method comprising:

(a) contacting (e.g., in vitro) a test compound with

(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide);

(e.g., and with a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell); and

(b) evaluating if the test compound modulates JMJD3 catalytic activity (e.g., as compared to a control). A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682

of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

In one aspect, the disclosure relates to a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. For example, with reference to (iii), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 1178-1682 of SEQ ID NO:1; and so forth. As another example, with reference to (iv), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 1178-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); and so forth. A sequence that is at least 85% identical to a sequence in (i)-(iv) can be, e.g., at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the recited sequence.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 (e.g., at residue 1178) and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is in soluble form.

In some embodiments, the polypeptide is in crystal form. In some embodiments, the crystal has the structure deposited in Protein Data Bank (PDB) ID: 2XUE.

In one embodiment, the polypeptide is suitable for crystallization, i.e., preferably the polypeptide fragment is crystallizable. Preferably, the crystals obtainable from the polypeptide described herein

are suitable for structure determination of the polypeptide using X-ray crystallography. Preferably, the crystals are radiation stable enough to permit more than 85% diffraction data completeness at resolution of preferably 3.5 Å or better to be collected upon exposure to monochromatic X-rays.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In another aspect, the disclosure relates to an isolated nucleic acid encoding an isolated polypeptide described herein.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In one aspect, the disclosure relates to a nucleic acid (e.g., an isolated nucleic acid), wherein the nucleic acid comprises a JMJD3 nucleic acid sequence, wherein the encoded polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 nucleic acid sequence consists of:

(i) a sequence at least 85% identical to nucleotides 3917-5422 of SEQ ID NO:2;

(ii) a sequence at least 85% identical to nucleotides 3917-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2;

(iii) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1115 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. When used for protein expression, the nucleic acid, e.g., with respect to (iii) and (iv), is in the proper reading frame to encode a polypeptide that possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)).

In some embodiments, the JMJD3 nucleic acid sequence consists of (i) a sequence at least 85% identical to nucleotides 3917-5422 of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to nucleotides 3917-5422 of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (ii) a sequence at least 85% identical to nucleotides 3917-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to nucleotides 3917-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of nucleotides 1133 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2.

In some embodiments, the JMJD3 nucleic acid sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1115 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the JMJD3 nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of nucleotides 1115 to 3916 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 nucleic acid sequence).

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 nucleic acid sequence) at 5' or 3' end. For example, the heterologous sequence can include a sequence that encodes a peptide used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a sequence encoding a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag, e.g., at its 3' end. In some embodiments, the nucleic acid comprises a sequence encoding a FLAG tag, e.g., at its 5' end. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag and a FLAG tag, e.g., at its 5' or 3' end. In some embodiments, the nucleic acid comprises a sequence encoding peptides used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the encoded polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the encoded polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a JMJD3 sequence (e.g., a polypeptide described herein), the method comprising
culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for identifying a compound which modulates (e.g., increases or decreases) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:

(a) constructing a computer model of a polypeptide comprising a JMJD3 sequence defined by the structure coordinates of the polypeptide, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

(b) selecting a test compound by a method selected from the group consisting of:

(i) assembling molecular fragments into the compound, (ii) selecting a compound from a small molecule database, and (iii) de novo ligand design of the compound;

(c) employing computational methods to perform a fitting program operation between computer models of the test compound and the polypeptide in order to provide an energy-minimized configuration of the compound with the polypeptide; and

(d) evaluating the results of the fitting operation to quantify the association between the test compound and the polypeptide computer model, thereby evaluating the ability of the test compound to associate with the polypeptide.

In some embodiments, the structure coordinates are deposited in Protein Data Bank (PDB) ID: 2XUE.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of

a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the method further comprises contacting the test compound with the polypeptide or a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide) (e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri-methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri-methylated, at K27)); the substrate can be naturally occurring in the host cell) (e.g., wherein the contacting occurs in conditions that are favorable for JMJD3 catalytic activity in the absence of a test compound); and

evaluating if the test compound modulates JMJD3 catalytic activity of the polypeptide (e.g., as compared to a control). A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

In another aspect, the disclosure relates to a method for identifying compounds which modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:

(a) contacting (e.g., in vitro) a test compound with

(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide)

(e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell); and

(b) analyzing the ability of the test compound to modulate JMJD3 catalytic activity (e.g., H3K27 demethylase activity) of the polypeptide. A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell).

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating (e.g., *in vitro*) a test compound for its ability to modulate (e.g., decrease or increase) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), the method comprising:
(a) incubating a test compound with
(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or
(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide),
(e.g., and with a JMJD3 substrate e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell)
wherein the incubating occurs in conditions that are favorable for JMJD3 catalytic activity in the absence of a test compound; and
(b) evaluating whether the test compound modulates JMJD3 catalytic activity, e.g., as compared to a control. A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some

embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating a test compound, the method comprising:
(a) contacting (e.g., in vitro) a test compound with
(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a JMJD3 sequence, wherein the polypeptide possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)); and wherein the JMJD3 sequence consists of:
(i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1;
(ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;
(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or
(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;
or
(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide);

(e.g., and with a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the residue corresponding to K27) (e.g., a peptide with the sequence ATKAARKSAPATGGVKKPHRYRPG (SEQ ID NO:5) (e.g., that is methylated, e.g., di- or tri- methylated, at the residue corresponding to K27); or histone 3 (e.g., that is methylated, e.g., di- or tri- methylated, at K27)); the substrate can be naturally occurring in the host cell); and

(b) evaluating if the test compound modulates JMJD3 catalytic activity (e.g., as compared to a control). A co-factor, such as Fe(II) and/or 2OG, can optionally be included (or may, e.g., be naturally occurring in the host cell). For example, the control can be the amount (e.g., percentage) of JMJD3 catalytic activity in the absence of a test compound under identical conditions.

In some embodiments, the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1181-1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the JMJD3 sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1180 (e.g., at residue 1178) of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the JMJD3 sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the JMJD3 sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage

site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in JMJD3 catalytic activity indicates that the test compound decreases (e.g., inhibits) JMJD3 catalytic activity.

In some embodiments, an increase in the JMJD3 catalytic activity indicates that the test compound increases (e.g., promotes) JMJD3 catalytic activity.

In some embodiments, the JMJD3 catalytic activity is determined by measuring demethylase activity (e.g., demethylation of di- or tri-methylated H3K27; e.g., the amount of mono-, di-, or tri-methylated H3K27).

In some embodiments, the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

In one aspect, the disclosure relates to a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a GATA-like domain sequence, wherein the polypeptide is capable of binding Zn++; and wherein the GATA-like domain sequence consists of:

(i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. For example, with reference to (iv), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 253-1682 of SEQ ID NO:1; 281-1682 of SEQ ID NO:1; residues 400-1682 of SEQ ID NO:1; residues 600-1682 of SEQ ID NO:1; residues 1000-1682 of SEQ ID NO:1; residues 1100-1682 of SEQ ID NO:1; residues 1140-1682 of SEQ ID NO:1; residues 1300-1682 of SEQ ID NO:1; residues 1400-1682 of SEQ ID NO:1; residues 1500-1682 of SEQ ID NO:1; residues 1540-1682 of SEQ ID NO:1; and so forth. As another example, with reference to (v), the JMJD3 sequence can consist of a sequence at least 85% identical to residues 247-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 253-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 281-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 400-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 600-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1000-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1100-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1140-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1300-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1400-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1500-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); residues 1540-1682 of SEQ ID NO:1 (with a deletion of residues 1637-1675); and so forth. A sequence that is at least 85% identical to a sequence in (i)-(v) can be, e.g., at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the recited sequence.

In some embodiments, the GATA-like domain sequence consists of (i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1543-1626 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the GATA-like domain sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the GATA-like domain sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, the polypeptide is in soluble form.

In some embodiments, the polypeptide is in crystal form. In some embodiments, the crystal has the structure deposited in Protein Data Bank (PDB) ID: 2XUE.

In one embodiment, the polypeptide is suitable for crystallization, i.e., preferably the polypeptide fragment is crystallizable. Preferably, the crystals obtainable from the polypeptide described herein are suitable for structure determination of the polypeptide using X-ray crystallography. Preferably, the crystals are radiation stable enough to permit more than 85% diffraction data completeness at resolution of preferably 3.5 Å or better to be collected upon exposure to monochromatic X-rays.

In another aspect, the disclosure relates to an isolated nucleic acid encoding an isolated polypeptide described herein.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a GATA-like domain (e.g., a polypeptide described herein), the method comprising
culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a GATA-like domain (e.g., a polypeptide described herein), the method comprising
culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In one aspect, the disclosure relates to the use of a polypeptide (e.g., isolated polypeptide) that contains the GATA-like domain described herein as a dominant negative protein to decrease (e.g., inhibit) JMJD3 activity. The polypeptide can comprise the GATA-like domain (e.g., residues 1543-1626 of SEQ ID NO:1, (or a sequence at least 85% identical thereto)) with an additional heterologous sequence, e.g., of about 10, about 20, about 30, about 40, or about 50 additional residues, e.g., at the amino or carboxy terminus of the GATA-like domain. The GATA-like domain is capable of binding Zn++.

In one aspect, the disclosure relates to a nucleic acid (e.g., an isolated nucleic acid), wherein the nucleic acid comprises a GATA-like domain nucleic acid sequence, wherein the encoded polypeptide is capable of binding Zn++; and wherein the GATA-like domain nucleic acid sequence consists of:

(i) a sequence at least 85% identical to nucleotides 5003-5254 of SEQ ID NO:2;

(ii) a sequence at least 85% identical to nucleotides 3797-5422 of SEQ ID NO:2;

(iii) a sequence at least 85% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2;

(iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 5002 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2; or

(v) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1115 to 5002 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. When used for protein expression, the nucleic acid, e.g., with respect to (iv) and (v), is in the proper reading frame to encode a polypeptide that possesses a catalytic activity of JMJD3 (e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)).

In some embodiments, the GATA-like domain nucleic acid sequence consists of (i) a sequence at least 85% identical to nucleotides 5003-5254 of SEQ ID NO:2. In some embodiments, the GATA-like domain nucleic acid sequence consists of a sequence 100% identical to nucleotides 5003-5254 of SEQ ID NO:2.

In some embodiments, the GATA-like domain nucleic acid sequence consists of (ii) a sequence at least 85% identical to nucleotides 3797-5422 of SEQ ID NO:2. In some embodiments, the GATA-like domain nucleic acid sequence consists of a sequence 100% identical to nucleotides 3797-5422 of SEQ ID NO:2.

In some embodiments, the GATA-like domain nucleic acid sequence consists of (iii) a sequence at least 85% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the GATA-like domain nucleic acid sequence consists of a sequence 100% identical to nucleotides 3797-5422 with a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the GATA-like domain nucleic acid sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of nucleotides 1133 to 5003 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2. In some embodiments, the GATA-like domain nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of nucleotides 1133 to 5003 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2.

In some embodiments, the GATA-like domain nucleic acid sequence consists of (v) a sequence at least 85% identical to a sequence that initiates at any of positions 1115 to 5003 of SEQ ID NO:2 and

terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2. In some embodiments, the GATA-like domain nucleic acid sequence consists of a sequence 100% identical to a sequence that initiates at any of positions 1115 to 5003 of SEQ ID NO:2 and terminates at nucleotide 5422 of SEQ ID NO:2, wherein the sequence comprises a deletion of nucleotides 5285-5401 (del 5285-5401) of SEQ ID NO:2.

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the GATA-like domain nucleic acid sequence).

In some embodiments, the nucleic acid comprises a heterologous sequence (e.g., non naturally occurring with the GATA-like domain nucleic acid sequence) at 5' or 3' end. For example, the heterologous sequence can include a sequence that encodes a peptide used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a sequence encoding a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag, e.g., at its 3' end. In some embodiments, the nucleic acid comprises a sequence encoding a FLAG tag, e.g., at its 5' end. In some embodiments, the nucleic acid comprises a sequence encoding a 6-His tag and a FLAG tag, e.g., at its 5' or 3' end. In some embodiments, the nucleic acid comprises a sequence encoding peptides used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the encoded polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In another aspect, the disclosure relates to a recombinant vector comprising the isolated nucleic acid.

In another aspect, the disclosure relates to a recombinant host cell comprising the isolated nucleic acid or the recombinant vector.

In another aspect, the disclosure relates to a method for making a polypeptide comprising a GATA-like domain (e.g., a polypeptide described herein), the method comprising
culturing the recombinant host cell under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to a method for making a polypeptide comprising a GATA-like domain (e.g., a polypeptide described herein), the method comprising

culturing a recombinant host cell containing a nucleic acid or recombinant vector that encodes a polypeptide that comprises a JMJD3 sequence under conditions permitting expression of the polypeptide from the nucleic acid or recombinant vector. The method can further comprise isolating the polypeptide from the host cell. The polypeptide can be isolated, for example, from a cell lysate. The isolating can be performed, e.g., by affinity chromatography (e.g., utilizing a tag (e.g., poly-His (e.g., 6-His), Flag, or c-myc tag, etc.).

In another aspect, the disclosure relates to an antibody directed against a polypeptide described herein (e.g., an epitope thereof). For example, the antibody can bind to the GATA-like domain.

In another aspect, the disclosure relates to a method for identifying compounds which modulate (e.g., increase or decrease) GATA-like domain Zn++ binding, the method comprising:
(a) contacting (e.g., in vitro) a test compound with
(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a GATA-like domain sequence, wherein the polypeptide is capable of binding Zn++; and wherein the GATA-like domain sequence consists of:

(i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or
(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide)
(e.g., and in the presence of Zn++; the Zn++ can be naturally occurring in the host cell); and
(b) analyzing the ability of the test compound to modulate Zn++ binding by the polypeptide.

In some embodiments, the GATA-like domain sequence consists of (i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1543-1626 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID

NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the GATA-like domain sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the GATA-like domain sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in Zn++ binding by the polypeptide indicates that the test compound decreases (e.g., inhibits) Zn++ binding.

In some embodiments, an increase in the Zn++ binding by the polypeptide indicates that the test compound increases (e.g., promotes) Zn++ binding.

In some embodiments, Zn++ binding is determined by measuring the level of free Zn++ (e.g., an increase (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound decreases Zn++ binding by the polypeptide; e.g., a decrease (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound increases Zn++ binding by the polypeptide).

In some embodiments, the level of free Zn++ is determined by a spectrophotometric assay.

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating (e.g., *in vitro*) a test compound for its ability to modulate (e.g., decrease or increase) GATA-like domain Zn++ binding, the method comprising:

(a) incubating a test compound with

(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a GATA-like domain sequence, wherein the polypeptide is capable of binding Zn++; and wherein the GATA-like domain sequence consists of:

(i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide)

(e.g., and in the presence of Zn++; the Zn++ can be naturally occurring in the host cell);

wherein the incubating occurs in conditions that are favorable for GATA-like domain Zn++ binding in the absence of a test compound; and

(b) evaluating whether the test compound modulates GATA-like domain Zn++ binding, e.g., as compared to a control. For example, the control can be the amount (e.g., percentage) of GATA-like domain Zn++ binding in the absence of a test compound under identical conditions.

In some embodiments, the GATA-like domain sequence consists of (i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1543-1626 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the GATA-like domain sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the GATA-like domain sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage

site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in Zn++ binding by the polypeptide indicates that the test compound decreases (e.g., inhibits) Zn++ binding.

In some embodiments, an increase in the JMJD3 catalytic activ Zn++ binding by the polypeptide indicates that the test compound increases (e.g., promotes) Zn++ binding.

In some embodiments, Zn++ binding is determined by measuring the level of free Zn++ (e.g., an increase (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound decreases Zn++ binding by the polypeptide; e.g., a decrease (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound increases Zn++ binding by the polypeptide).

In some embodiments, the level of free Zn++ is determined by a spectrophotometric assay.

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

A method of evaluating a test compound, the method comprising:
(a) contacting (e.g., in vitro) a test compound with

(1) a polypeptide (e.g., an isolated polypeptide), wherein the polypeptide comprises a GATA-like domain sequence, wherein the polypeptide is capable of binding Zn++; and wherein the GATA-like domain sequence consists of:

(i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

or

(2) a recombinant host cell comprising the polypeptide (e.g., a recombinant host cell described herein; e.g., containing a nucleic acid encoding the polypeptide)

(e.g., and in the presence of Zn++; the Zn++ can be naturally occurring in the host cell); and

(b) evaluating if the test compound modulates GATA-like domain Zn++ binding (e.g., as compared to a control). For example, the control can be the amount (e.g., percentage) of GATA-like domain Zn++ binding in the absence of a test compound under identical conditions.

In some embodiments, the GATA-like domain sequence consists of (i) a sequence at least 85% identical to residues 1543-1626 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1543-1626 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to identical to a sequence that initiates at any of residues 253 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

In some embodiments, the GATA-like domain sequence consists of (v) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 15430 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1. In some embodiments, the GATA-like domain sequence consists of a sequence 100% identical to a sequence that initiates at any of residues 247 to 1543 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., a sequence that does not naturally occur with the GATA-like domain sequence).

In some embodiments, the polypeptide comprises a heterologous sequence (e.g., non naturally occurring with the GATA-like domain sequence) at its amino or carboxyl terminus. For example, the heterologous sequence can include a sequence used to detect or purify the polypeptide, e.g., a poly His (e.g., 6-His), FLAG, or c-myc tag, and/or can include a cleavage site, e.g., a protease cleavage site, e.g., a thrombin, Factor Xa, PreScission protease, or TEV protease cleavage site. In some embodiments, the polypeptide comprises a 6-His tag, e.g., at its carboxy terminus. In some embodiments, the polypeptide comprises a FLAG tag, e.g., at its amino terminus. In some embodiments, the polypeptide comprises a 6-His tag and a FLAG tag, e.g., at its amino or carboxy terminus. In some embodiments, the polypeptide comprises a sequence used to detect or purify the polypeptide and a cleavage site, e.g., a 6-His tag and a TEV protease cleavage site.

In some embodiments, the polypeptide comprises Cys (cysteine) at residues 1575, 1578, 1602, and 1605 of SEQ ID NO:1.

In some embodiments, a decrease in Zn++ binding by the polypeptide indicates that the test compound decreases (e.g., inhibits) Zn++ binding.

In some embodiments, an increase in Zn++ binding by the polypeptide indicates that the test compound increases (e.g., promotes) Zn++ binding.

In some embodiments, Zn++ binding is determined by measuring the level of free Zn++ (e.g., an increase (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound decreases Zn++ binding by the polypeptide; e.g., a decrease (e.g., as compared to a negative control) in free Zn++ in the presence of a test compound indicates that the compound increases Zn++ binding by the polypeptide).

In some embodiments, the level of free Zn++ is determined by a spectrophotometric assay.

In some embodiments, the test compound is comprised in a compound library/ members of a compound library are evaluated.

In some embodiments, the test compound comprises a nucleic acid, a protein, or a small molecule.

In some embodiments, the test compound comprises RNAi.

In some embodiments, the RNAi is selected from the group consisting of: miRNA, siRNA, esiRNA, and shRNA.

In some embodiments, the test compound comprises a chemical compound.

In one aspect, the disclosure relates to a polypeptide (e.g., isolated polypeptide) that possesses JMJD3 catalytic activity (e.g., H3K27 demethylase activity), e.g., wherein the polypeptide contains residues 1141-1682 of SEQ ID NO:1 (or a sequence at least 85% identical thereto) or residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1 (or a sequence at least 85% identical thereto), but not the full length JMJD3 sequence. Nucleic acids encoding the same are also described.

In one aspect, the disclosure relates to a polypeptide (e.g., isolated polypeptide) that possesses JMJD3 catalytic activity (e.g., H3K27 demethylase activity), e.g., wherein the polypeptide contains residues 1181-1682 of SEQ ID NO:1 (or a sequence at least 85% identical thereto) or residues 1181-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1 (or a sequence at least 85% identical thereto), but not the full length JMJD3 sequence. Nucleic acids encoding the same are also described.

In one aspect, the disclosure relates to a polypeptide (e.g., isolated polypeptide) that contains the GATA-like domain and binds Zn++, e.g., wherein the polypeptide contains residues 1543-1626 of SEQ ID NO:1 (or a sequence at least 85% identical thereto), but not the full length JMJD3 sequence. Nucleic acids encoding the same are also described.

In one embodiment, a polypeptide described herein is suitable for crystallization, e.g., the polypeptide fragment is crystallizable. Further, the crystals obtainable from a polypeptide described herein are suitable for structure determination of the polypeptide using X-ray crystallography. In some embodiments, the crystals are radiation stable enough to permit more than 85% diffraction data completeness at resolution of e.g., 3.5 Å or better to be collected upon exposure to monochromatic X-rays.

Brief Description of the Drawings

FIG. 1 is a diagram of the overall architecture of JMJD3 showing residues 1178 to 1488 of the catalytic domain (top domain of diagram- residues 1141-1488) and the GATA-like insert (in darker

shading) (residues 1543-1626) within the four helical domain (bottom domain of diagram-residues1489-1636) plus residues 1676 and 1677 after the deletion 1637-1675 (del 1637-1675).

FIG. 2 is a series of six panels showing the co-ordination of 2-OG in active sites of JMJD3, PHF8 (PDB ID: 3K3O), JMJD2A (PDB ID: 2GP5), and FBX11 (PDB ID: 2YU1). Also shown is PDCA in the active site of Jmjd2A (PDB ID: 2VD7) and finally PDCA in Jmjd2A (PDB ID: 2VD7) superposed on 2-OG in JMJD3.

FIG. 3 is a series of three structures of C-terminal helical domains of PHF8 (center), FBX11 (right) and JMJD3 (left) with the Zn-binding GATAL insert.

FIG. 4 is a diagram showing the topology of the zinc-binding domain in JMJD3. The two strands that are "added" to the treble clef fold are highlighted in darker shading on the left side of the diagram.

FIG. 5 is a diagram showing JMJD3. The JmjC domain is at the left; helical domain at the bottom right; and GATAL at the top right of the structure. The FOG fragment (labelled) is positioned consistent with its corresponding interactions with the GATA1 protein in the GATA-1/FOG complex (PDB ID: 1Y0J).

FIG. 6 is a bar graph showing the quantification of the respective H3K27 methyl states on an Arrayscan and demonstrates JmjD3t (JMJD3 1141-1682 (del. 1637-1675)) substrate preference for the tri- and di- methyl valencies. All data is mean of 3 independent experiments.

FIG. 7 is a series of three diagrams: (i) GATA-1 (right) and DNA (left) showing the long C-terminal tail of GATA-1 that binds to the DNA minor groove; (ii) The GATAL of JMJD3 in the same orientation to highlight that instead of a long C-terminal tail, JMJD3 has a C-terminal region that forms one of the additional strands absent in GATA-1; (iii) an overlay of the two GATA domains in the presence of DNA is shown. The GATA-1 domain is from PDB ID: 2DFV.

Detailed Description of the Invention

Studies have identified JmjC domain-containing proteins as histone demethylases that mediate the reversal of methylation at histone H3K4, H3K9 and H3K36. The JMJD3 family of proteins is a subfamily of JmjC domain-containing proteins that has been shown to demethylate H3K27 (e.g., K27 on human histone 3). The polypeptides, nucleic acids, crystal structures, and assays described herein can be utilized to screen for and identify compounds that modulate (e.g., increase or decrease) JMJD3 catalytic activity. Polypeptides, nucleic acids, crystal structures, and assays described herein can also be used to screen for and identify compounds that modulate (e.g., increase or decrease) Zn++ binding by the JMJD3 GATA-like domain.

JMJD3

The nucleotide and amino acid sequences for human JMJD3 are provided below. All sequence numbers are based on the numbering of residues taken from NCBI reference sequence entries NM_001080424 (nucleotide sequence) and NP_001073893 (amino acid sequence).

The human JMJD3 amino acid sequence is as follows. The amino acid sequence for the GATA-like zinc binding domain is highlighted in bold and underlined in the sequence below (taken from NP_001073893).

```
   1 mhravdppga raareafalg glscagawss cpphppprsa wlpggrcsas igqpplpapl
  61 ppshgsssgh pskpyyapga ptprplhgkl eslhgcvqal lrepaqpglw eqlgqlyese
 121 hdseeatrcy hsalryggsf aelgprigrl qqaqlwnfht gscqhrakvl ppleqvwnll
 181 hlehkrnyga krggppvkra aeppvvqpvp paalsgpsge eglspggkrr rgcnseqtgl
 241 ppglplpppp lppppppppp ppppplpglat sppfqltkpg lwstlhgdaw gperkgsapp
 301 erqeqrhslp hpypypapay tahppghrlv paappgpgpr ppgaeshgcl patrppgsdl
 361 resrvqrsrm dssvspaatt acvpyapsrp pglpgtttss sssssntgl rgvepnpgip
 421 gadhyqtpal evshhgrlgp sahssrkpfl gapaatphls lppgpssppp ppcprllrpp
 481 pppawlkgpa craaredgei leelffgteg pprpappplp hregflgppa srfsvgtqds
 541 htpptpptpt tsssnsnsgs hssspagpvs fppppylars idplprppsp aqnpqdpplv
 601 pltlalppap psschqntsg sfrrpesprp rvsfpktpev gpgpppgpls kapqpvppgv
 661 gelpargprl fdfpptpled qfeepaefki lpdglanimk mldesirkee eqqqheagva
 721 pqpplkepfa slqspfptdt aptttapava vttttttttt ttatqeeekk pppalppppp
 781 lakfpppsqp qpppppppsp asllkslasv legqkycyrg tgaavstrpg plpttqyspg
 841 ppsgatalpp tsaapsaqgs pqpsassssq fstsggpwar erragеepvp gpmtptqppp
 901 plslpparse sevleeisra cetlvervgr satdpadpvd taepadsgte rllppaqake
 961 eaggvaavsg sckrrqkehq kehrrhrrac kdsvgrrpre grakakakvp keksrrvlgn
1021 ldlqseeiqg reksrpdlgg askakpptap appsapapsa qptppsasvp gkkareeapg
1081 ppgvsradml klrslsegpp kelkirlikv esgdketfia seveerrlrm adltishcaa
1141 dvvrasrnak vkgkfresyl spaqsvkpki nteeklprek lnpptpsiyl eskrdafspv
1201 llqfctdprn pitvirglag slrlnlglfs tktlveasge htvevrtqvq qpsdenwdlt
1261 gtrqiwpces srshttiaky aqyqassfqe slqeekesed eeseepdstt gtppssapdp
1321 knhhiikfgt nidlsdakrw kpqlqellkl pafmrvtstg nmlshvghti lgmntvqlym
1381 kvpgsrtpgh qennnfcsvn inigpgdcew favhehywet isafcdrhgv dyltgswwpi
1441 lddlyasnip vyrfvqrpgd lvwinagtvh wvqatgwcnn iawnvgplta yqyqlalery
1501 ewnevknvks ivpmihvswn vartvkisdp dlfkmikfcl lqsmkhcqvq reslvragkk
1561 iayggrvkde payycnecdv evfnilfvts engsrntylv hcegcarrrs aglqgvvvle
1621 qyrteelaqa ydaftlvrar rargqrrral gqaagtgfgs paapfpeppp afspqapast
1681 sr
```

(SEQ ID NO:1)

The human JMJD3 nucleotide sequence is provided below. The nucleotide sequence for the zinc binding domain is highlighted in bold and underlined in the sequence below (taken from NM_001080424). Also marked in this sequence are the methionine (atg) that is the start of the JMJD3 sequence and the stop codon (tga) for JMJD3, which are both highlighted in bold and underlined and italicized.

```
   1 ggcaacatgc cagccccgta gcactgccca ccccacccac tgtggtctgt tgtaccccac
  61 tgctggggtg gtggttccaa tgagacaggg cacaccaaac tccatctggc tgttactgag
 121 gcggagacac gggtgatgat tggctttctg gggagagagg aagtcctgtg attggccaga
 181 tctctggagc ttgccgacgc ggtgtgagga cgctcccacg gaggccggaa ttggctgtga
 241 aaggactgag gcagccatct gggggtagcg ggcactctta tcagagcggc tggagccgga
 301 ccatcgtccc agagagctgg ggcaggggcc cgtgcccaat ctccagggct cctggggcca
 361 ctgctgacct ggctggatgc atcgggcagt ggaccctcca ggggcccgcg ctgcacggga
 421 agcctttgcc cttgggggcc tgagctgtgc tggggcctgg agctcctgcc cgcctcatcc
```

```
 481 ccctcctcgt agcgcatggc tgcctggagg cagatgctca gccagcattg ggcagccccc
 541 gcttcctgct cccctacccc cttcacatgg cagtagttct gggcacccca gcaaaccata
 601 ttatgctcca ggggcgccca ctccaagacc cctccatggg aagctggaat ccctgcatgg
 661 ctgtgtgcag gcattgctcc gggagccagc ccagccaggg ctttgggaac agcttgggca
 721 actgtacgag tcagagcacg atagtgagga ggccacacgc tgctaccaca gcgcccttcg
 781 atacggagga agcttcgctg agctggggcc ccgcattggc cgactgcagc aggcccagct
 841 ctggaacttt catactggct cctgccagca ccgagccaag gtcctgcccc cactggagca
 901 agtgtggaac ttgctacacc ttgagcacaa acggaactat ggagccaagc ggggaggtcc
 961 cccggtgaag cgagctgctg aacccccagt ggtgcagcct gtgcctcctg cagcactctc
1021 aggcccctca ggggaggagg gcctcagccc tggaggcaag cgaaggagag gctgcaactc
1081 tgaacagact ggccttcccc cagggctgcc actgcctcca ccaccattac caccaccacc
1141 accaccacca ccaccaccac caccacccct gcctggcctg gctaccagcc ccccatttca
1201 gctaaccaag ccagggctgt ggagtaccct gcatggagat gcctggggcc cagagcgcaa
1261 gggttcagca cccccagagc gccaggagca gcggcactcg ctgcctcacc catatccata
1321 cccagctcca gcgtacaccg cgcacccccc tggccaccgg ctggtcccgg ctgctccccc
1381 aggcccaggc ccccgccccc caggagcaga gagccatggc tgcctgcctg ccacccgtcc
1441 ccccggaagt gaccttagag agagcagagt tcagaggtcg cggatggact ccagcgtttc
1501 accagcagca accaccgcct gcgtgcctta cgcccccttcc cggccccctg gcctccccgg
1561 caccaccacc agcagcagca gtagcagcag cagcaacact ggtctccggg gcgtggagcc
1621 gaacccaggc attcccggcg ctgaccatta ccaaactccc gcgctggagg tctctcacca
1681 tggccgcctg gggccctcgg cacacagcag tcggaaaccg ttcttggggg ctcccgctgc
1741 cactccccac ctatccctgc cacctggacc ttcctcaccc cctccacccc cctgtccccg
1801 cctcttacgc cccccaccac cccctgcctg gttgaagggt ccggcctgcc gggcagcccg
1861 agaggatgga gagatcttag aagagctctt ctttgggact gagggacccc cccgccctgc
1921 cccaccaccc ctcccccatc gcgagggctt cttggggcct ccggcctccc gcttttctgt
1981 gggcactcag gattctcaca cccctcccac tcccccaacc ccaaccacca gcagtagcaa
2041 cagcaacagt ggcagccaca gcagcagccc tgctgggcct gtgtcctttc ccccaccacc
2101 ctatctggcc agaagtatag acccccttcc ccggcctccc agcccagcac agaaccccca
2161 ggacccacct cttgtacccc tgactcttgc cctgcctcca gcccctcctt cctcctgcca
2221 ccaaaatacc tcaggaagct tcaggcgccc ggagagcccc cggcccaggg tctccttccc
2281 aaagaccccc gaggtggggc cggggccacc cccaggcccc ctgagtaaag ccccccagcc
2341 tgtgccgccc ggggttgggg agctgcctgc ccgaggccct cgactctttg attttccccc
2401 cactccgctg gaggaccagt ttgaggagcc agccgaattc aagatcctac ctgatgggct
2461 ggccaacatc atgaagatgc tggacgaatc cattcgcaag gaagaggaac agcaacaaca
2521 cgaagcaggc gtggcccccc aacccccgct gaaggagccc tttgcatctc tgcagtctcc
2581 tttccccacc gacacagccc ccaccactac tgctcctgct gtcgccgtca ccaccaccac
2641 caccaccacc accaccacca cggccaccca ggaagaggag aagaagccac caccagccct
2701 accaccacca ccgcctctag ccaagttccc tccaccctct cagccacagc caccaccacc
2761 cccacccccc agcccggcca gcctgctcaa atccttggcc tccgtgctgg agggacaaaa
2821 gtactgttat cgggggactg gagcagctgt ttccacccgg cctgggccct tgcccaccac
2881 tcagtattcc cctggccccc catcaggtgc taccgccctg ccgcccacct cagcggcccc
2941 tagcgcccag ggctccccac agccctctgc ttcctcgtca tctcagttct ctacctcagg
3001 cgggccctgg gcccgggagc gcagggcggg cgaagagcca gtcccgggcc ccatgacccc
3061 cacccaaccg cccccacccc tatctctgcc ccctgctcgc tctgagtctg aggtgctaga
3121 agagatcagc cgggcttgcg agacccttgt ggagcgggtg ggccggagtg ccactgaccc
3181 agccgaccca gtggacacag cagagccagc ggacagtggg actgagcgac tgctgccccc
3241 cgcacaggcc aaggaggagg ctggcggggt ggcggcagtg tcaggcagct gtaagcggcg
3301 acagaaggag catcagaagg agcatcggcg gcacaggcgg gcctgtaagg acagtgtggg
3361 tcgtcggccc cgtgagggca gggcaaaggc caaggccaag gtccccaaag aaaagagccg
3421 ccgggtgctg gggaacctgg acctgcagag cgaggagatc cagggtcgtg agaagtcccg
3481 gcccgatctt ggcggggcct ccaaggccaa gccacccaca gctccagccc ctccatcagc
3541 tcctgcacct tctgcccagc ccacaccccc gtcagcctct gtccctggaa agaaggctcg
3601 ggaggaagcc ccagggccac cgggtgtcag ccgggccgac atgctgaagc tgcgctcact
3661 tagtgagggg ccccccaagg agctgaagat ccggctcatc aaggtagaga gtggtgacaa
3721 ggagaccttt atcgcctctg aggtggaaga gcggcggctg cgcatggcag acctcaccat
3781 cagccactgt gctgctgacg tcgtgcgcgc cagcaggaat gccaaggtga aagggaagtt
3841 tcgagagtcc tacctttccc ctgcccagtc tgtgaaaccg aagatcaaca ctgaggagaa
3901 gctgccccgg gaaaaactca acccccctac acccagcatc tatctggaga gcaaacggga
3961 tgccttctca cctgtcctgc tgcagttctg tacagaccct cgaaatccca tcacagtgat
4021 ccggggcctg gcgggctccc tgcggctcaa cttgggcctc ttctccacca agaccctggt
4081 ggaagcgagt ggcgaacaca ccgtggaagt tcgcacccag gtgcagcagc cctcagatga
4141 gaactgggat ctgacaggca ctcggcagat ctggccttgt gagagctccc gttcccacac
4201 caccattgcc aagtacgcac agtaccaggc ctcatccttc caggagtctc tgcaggagga
4261 gaaggagagt gaggatgagg agtcagagga gccagacagc accactggaa cccctcctag
4321 cagcgcacca gacccgaaga accatcacat catcaagttt ggcaccaaca tcgacttgtc
4381 tgatgctaag cggtggaagc cccagctgca ggagctgctg aagctgcccg ccttcatgcg
4441 ggtaacatcc acgggcaaca tgctgagcca cgtgggccac accatcctgg gcatgaacac
4501 ggtgcagctg tacatgaagg tgcccggcag ccgaacgcca ggccaccagg agaataacaa
```

```
4561 cttctgctcc gtcaacatca acattggccc aggcgactgc gagtggttcg cggtgcacga
4621 gcactactgg gagaccatca gcgctttctg tgatcggcac ggcgtggact acttgacggg
4681 ttcctggtgg ccaatcctgg atgatctcta tgcatccaat attcctgtgt accgcttcgt
4741 gcagcgaccc ggagacctcg tgtggattaa tgcggggact gtgcactggg tgcaggccac
4801 cggctggtgc aacaacattg cctggaacgt ggggcccctc accgcctatc agtaccagct
4861 ggccctggaa cgatacgagt ggaatgaggt gaagaacgtc aaatccatcg tgcccatgat
4921 tcacgtgtca tggaacgtgg ctcgcacggt caaaatcagc gaccccgact tgttcaagat
4981 gatcaagttc tgcctgctgc agtccatgaa gcactgccag gtgcaacgcg agagcctggt
5041 gcgggcaggg aagaaaatcg cttaccaggg ccgtgtcaag gacgagccag cctactactg
5101 caacgagtgc gatgtggagg tgtttaacat cctgttcgtg acaagtgaga atggcagccg
5161 caacacgtac ctggtacact gcgagggctg tgcccggcgc cgcagcgcag gcctgcaggg
5221 cgtggtggtg ctggagcagt accgcactga ggagctggct caggcctacg acgccttcac
5281 gctggtgagg gcccggcggg cgcgcgggca gcggaggagg gcactggggc aggctgcagg
5341 gacgggcttc gggagcccgg ccgcgccttt ccctgagccc ccgccggctt tctccccccа
5401 ggccccagcc agcacgtcgc gatgaggccg gacgccccgc ccgcctgcct gcccgcgcaa
5461 ggcgccgcgg ggccaccagc acatgcctgg gctggaccta ggtcccgcct gtggccgaga
5521 agggggtcgg gcccagccct tccaccccat tggcagctcc cctcacttaa tttattaaga
5581 aaaacttttt ttttttttt agcaaatatg aggaaaaaag gaaaaaaaat gggagacggg
5641 ggaggggggct ggcagcccct cgcccaccag cgcctcccct caccgacttt ggccttttta
5701 gcaacagaca caaggaccag gctccggcgg cggcgggggt cacatacggg ttccctcacc
5761 ctgccagccg cccgcccgcc cggcgcagat gcacgcggct cgtgtatgta catagacgtt
5821 acggcagccg aggtttttaa tgagattctt tctatgggct ttacccctcc cccggaacct
5881 cctttttac ttccaatgct agctgtgacc cctgtacatg tctctttatt cacttggtta
5941 tgatttgtat tttttgttct tttcttgttt ttttgttttt aatttataac agtcccactc
6001 acctctattt attcattttt gggaaaaccc gacctcccac acccccaagc catcctgccc
6061 gcccctccag ggaccgcccg tcgccgggct ctccccgcgc cccagtgtgt gtccgggccc
6121 ggcccgaccg tctccacccg tccgcccgcg gctccagccg ggttctcatg gtgctcaaac
6181 ccgctccccc cccctacgtc ctgcactttc tcggaccagt ccccccactc ccgacccgac
6241 cccagcccca cctgagggtg agcaactcct gtactgtagg ggaagaagtg ggaactgaaa
6301 tggtattttg taaaaaaaat aaataaaata aaaaaattaa aggttttaaa gaaagaacta
6361 tgaggaaaag gaaccccgtc cttcccagcc ccggccaact ttaaaaaaca cagaccttca
6421 cccccacccc cttttctttt taagtgtgaa acaacccagg gccagggcct cactggggca
6481 gggacacccc ggggtgagtt tctctggggc tttattttcg ttttgttggt tgttttttct
6541 ccacgctggg gctgcggagg ggtgggggggt ttacagtccc gcaccctcgc actgcactgt
6601 ctctctgccc caggggcaga ggggtcttcc caaccctacc cctattttcg gtgatttttg
6661 tgtgagaata ttaatattaa aaataaacgg agaaaaaaaa tcct
```

(SEQ ID NO:2)

JMJD3 possesses catalytic (e.g., demethylase) activity. For example, JMJD3 can demethylate both tri- and dimethylated H3K27 (H3K27 demethylase activity) (e.g., K27 on human histone 3). The polypeptides described herein contain a JMJD3 catalytic activity (e.g., H3K27 demethylase activity) and can contain amino acids (residues) 1141-1682 of the sequence shown in NP_001073893 or SEQ ID NO:1. The polypeptides can be encoded by sequences containing nucleotides 3797-5422 of the sequence shown in NM_001080424 or SEQ ID NO:2. As another example, the polypeptides described herein contain a JMJD3 catalytic activity (e.g., H3K27 demethylase activity) and can contain amino acids 1141-1682 (residues) and a deletion of amino acids (residues) 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1. The polypeptides can be encoded by sequences containing nucleotides 3797-5422 (del 5285-5401) of the sequence shown in NM_001080424 or SEQ ID NO:2. As further examples, the polypeptides described herein contain a JMJD3 catalytic activity (e.g., H3K27 demethylase activity) and can contain amino acids (residues) 1181-1682 of the sequence shown in NP_001073893 or SEQ ID NO:1. The polypeptides can be encoded by sequences containing nucleotides 3917-5422 of the sequence shown in NM_001080424 or SEQ ID NO:2; the polypeptides described herein contain a JMJD3 catalytic activity (e.g., H3K27 demethylase activity) and can contain amino acids 1181-1682 (residues) and a deletion of amino

acids (residues) 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1. The polypeptides can be encoded by sequences containing nucleotides 3917-5422 (del 5285-5401) of the sequence shown in NM_001080424 or SEQ ID NO:2.

JMJD3 binds to 2-oxo-glutarate (2OG). The polypeptides described herein can bind to 2OG and can contain amino acids 1141-1682 or amino acids 1141-1682 with a deletion of amino acids 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1. As other examples, the polypeptides described herein can bind to 2OG and can contain amino acids 1181-1682 or amino acids 1181-1682 with a deletion of amino acids 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1.

JMJD3 binds iron (Fe(II)). The polypeptides described herein can bind iron and can contain amino acids 1141-1682 or amino acids 1141-1682 with a deletion of amino acids 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1. As additional examples, the polypeptides described herein can bind iron and can contain amino acids 1181-1682 or amino acids 1181-1682 with a deletion of amino acids 1637-1675 (del 1637-1675) of the sequence shown in NP_001073893 or SEQ ID NO:1.

In some embodiments, such as for crystallography studies, JMJD3 (and JMJD3 polypeptides described herein) can bind to other divalent ions, such as Co(II) and Ni(II), in addition to Fe(II).

The catalytic site containing beta helix contains amino acids 1181-1488 of the sequence shown in NP_001073893 or SEQ ID NO:1. The catalytic site containing beta helix can be encoded by nucleotides 3917-4840 of the sequence shown in NM_001080424 or SEQ ID NO:2. Polypeptides comprising this catalytic site beta helix (or a sequence at least 85% identical thereto) are part of the disclosure, and can be used in the methods described herein for identifying compounds which modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity).

Polypeptides containing the JMJD3 GATA-like domain are also described herein. The polypeptides contain amino acids 1543-1626 of the sequence shown in NP_001073893 or SEQ ID NO:1. The polypeptides can bind Zn++. The conserved zinc-binding cysteine motif within the GATA-like domain is CxxC...CxxC and contains Cys1575, Cys 1578, Cys1602 and Cys1605. The GATA-like domain can be encoded by nucleotides 5003-5254 of the sequence shown in NM_001080424 or SEQ ID NO:2.

The GATA-like domain structure consists of a single treble-clef zinc finger fold elaborated with two extra strands. The structure is similar to the zinc binding domain in the erythroid transcription factor GATA-1.

Calculations of "homology" or "sequence identity" between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The optimal alignment is determined as the best score using the GAP program in the GCG software package with a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences.

In one embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, or at least 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the length of the reference sequence. For example, the reference sequence may be the length of a sequence described herein (e.g., 100% of the length of a JMJD3 sequence or GATA-like domain sequence described herein).

It is understood that the polypeptides may comprise additional amino acids not derived from JMJD3, e.g., tags, enzyme cleavage sites, etc. Such additional amino acids are not considered in an alignment used to determine percent identity, i.e., the additional amino acids are excluded from the calculation of the percent identity. In one embodiment, the percent identity is obtained when aligning the sequence of the polypeptide at least over a length of 375, 400, 425, 450, 475, 500, 510, 525, or 540 amino acids for the polypeptide that possesses JMJD3 catalytic activity; or at least over a length of 60, 65, 70, 75, 80, or 83 amino acids for the polypeptide that possesses the GATA-like domain.

As used herein, the term "substantially identical" (or "substantially homologous") is used herein to refer to a first amino acid or nucleic acid (e.g., nucleotide) sequence that contains a sufficient number of identical or equivalent (e.g., with a similar side chain, e.g., conserved amino acid substitutions) amino acid residues or nucleotides to a second amino acid or nucleic acid sequence such that the first and second amino acid or nucleic acid sequences have (or encode proteins having) similar activities, e.g., a catalytic activity (e.g., H3K27 demethylase activity), a binding preference or specificity, or other activity. For example, in the case of H3K27 demethylases, the second polypeptide encoded by or containing the sequence can have the same specificity and/or activity, and/or shares at least one substrate specificity and/or has at least 50%, at least 25%, or at least 10% of the activity relative to the same substrate.

Sequences similar or homologous (e.g., have at least about 85% sequence identity) to the sequences (amino acid and/or nucleotide sequences) disclosed herein are also part of this disclosure. In some embodiments, the sequence identity can be 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher.

In addition, substantial identity exists when the nucleic acid sequences hybridize under selective hybridization conditions (e.g., highly stringent hybridization conditions), to the complement of the other sequence. The nucleic acids may be present in whole cells, in a cell lysate, or in a partially purified or substantially pure form. As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: (1) low stringency hybridization conditions in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions); (2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; (3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; and (4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions and the ones that should be used unless otherwise specified. The disclosure includes nucleic acids that hybridize with low, medium, high, or very high stringency to a nucleic acid described herein or to a complement thereof, e.g., nucleic acids encoding a binding protein described herein. The nucleic acids can be the same length or within 30, 20, or 10% of the length of the reference nucleic acid. The nucleic acid can correspond to a sequence described herein.

A polypeptide of the disclosure may have mutations (e.g., at least one, two, three, four, five, six, seven, eight, nine, or ten, and/or less than 15, 10, 5, or 3) relative to a sequence described herein (e.g., conservative or non-essential amino acid substitutions), which do not have a substantial effect on polypeptide function (e.g., the polypeptide has at least 50%, 60%, 70%, 80%, 90%, 92%, 95%, 97%, 98%, or 100% of the activity of the reference polypeptide). Whether or not a particular substitution will be tolerated, i.e., will not adversely affect biological properties, such as catalytic activity can be predicted, e.g., by evaluating whether the mutation is conservative or by the method of Bowie, et al. (1990) Science 247:1306-1310.

In some embodiments, the polypeptide comprises amino acid substitutions, insertions, or deletions, e.g., less than 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 substitutions, insertions, or deletions, as compared to the sequence of amino acids 1141-1682 of SEQ ID NO:1, and wherein the polypeptide possesses a catalytic activity of JMJD3, e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of

histone 3 (H3)). The His at position 1390 (with reference to the full length JMJD3 sequence of SEQ ID NO:1) is needed for JMJD3 catalytic activity. Mutation of this residue to Ala (H1390A) abolishes catalytic activity.

In some embodiments, the polypeptide comprises a sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the sequence of amino acids 1141-1682 (del 1637-1675) of SEQ ID NO:1, and wherein the polypeptide possesses a catalytic activity of JMJD3, e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)). The His at position 1390 (with reference to the full length JMJD3 sequence of SEQ ID NO:1) is needed for JMJD3 catalytic activity. Mutation of this residue to Ala (H1390A) abolishes catalytic activity.

In some embodiments, the polypeptide comprises amino acid substitutions, insertions, or deletions, e.g., less than 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 substitutions, insertions, or deletions, as compared to the sequence of amino acids 1181-1682 of SEQ ID NO:1, and wherein the polypeptide possesses a catalytic activity of JMJD3, e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)). The His at position 1390 (with reference to the full length JMJD3 sequence of SEQ ID NO:1) is needed for JMJD3 catalytic activity. Mutation of this residue to Ala (H1390A) abolishes catalytic activity.

In some embodiments, the polypeptide comprises a sequence at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the sequence of amino acids 1181-1682 (del 1637-1675) of SEQ ID NO:1, and wherein the polypeptide possesses a catalytic activity of JMJD3, e.g., the polypeptide is capable of demethylating H3K27 (lysine 27 (K27) of histone 3 (H3)). The His at position 1390 (with reference to the full length JMJD3 sequence of SEQ ID NO:1) is needed for JMJD3 catalytic activity. Mutation of this residue to Ala (H1390A) abolishes catalytic activity.

In some embodiments, the polypeptide comprises amino acid substitutions, insertions, or deletions, e.g., less than 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 substitutions, insertions, or deletions, as compared to the sequence of amino acids 1543-1626 of SEQ ID NO:1, and wherein the polypeptide is capable of binding Zn++.

A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). It is possible for amino acid residues of a polypeptide described herein to include one or more conservative substitutions. In some embodiments, one or more of the

amino acids that line the active site of JMJD3 (identified above) are kept invariant or replaced by only conservative substitutions.

A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of the polypeptide, without abolishing or more preferably, without substantially altering a biological activity, whereas changing an "essential" amino acid residue results in a substantial loss of activity.

For example, His1390, Glu1392, and His 1470 (with reference to the full length JMJD3 sequence) form a triad that binds iron. The His at position 1390 is needed for JMJD3 catalytic activity and is an essential amino acid. Mutation of this residue to Ala (H1390A) abolishes catalytic activity. Glu1392 and His1470 are also needed for JMJD3 catalytic activity and are essential amino acids. Mutation of either of these residues is expected to abolish catalytic activity. However, polypeptides lacking catalytic activity may contain a GATA-like domain (described herein) and be able to bind to Zn++. As a result, these polypeptides are useful, e.g., in assays for evaluating a test compound for its ability to modulate Zn++ binding by the GATA-like domain.

As another example, Cys1575, Cys 1578, Cys1602 and Cys1605 are essential amino acids for the GATA-like domain.

An isolated polypeptide refers to a preparation at least about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of which is made up of the polypeptide. For example, the isolated polypeptide can make up at least about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of the preparation.

An isolated nucleic acid or polynucleotide refers to a preparation at least about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of which is made up of the nucleic acid or polynucleotide. For example, the isolated nucleic acid or polynucleotide can make up at least about 50%, about 60%, about 70%, about 80%, about 90%, about 95% or more of the preparation.

In some embodiments, the polypeptide according to the present disclosure is purified to an extent to be suitable for crystallization, for example, it is 85% to 100%, 90% to 100%, or 95% to 100% pure.

In another embodiment, a polypeptide described herein, e.g., a polypeptide containing a JMJD3 sequence, is capable of binding to Fe(II).

In another embodiment, a polypeptide described herein, e.g., a polypeptide containing a JMJD3 sequence, is capable of binding to 2-oxo-glutarate (2OG).

In another embodiment, a polypeptide described herein, e.g., a polypeptide containing the GATA-like domain, is capable of binding Zn++.

In some embodiments, a polypeptide described herein (e e.g., a polypeptide containing a JMJD3 sequence) is not inhibited by pyridine 2,4 dicarboxylic acid (PDCA).

In one embodiment, the polypeptide consists of amino acids 1141-1682 (del 1637-1675) of SEQ ID NO:1 and has the structure defined by the structure coordinates deposited on October 19, 2010 in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) and assigned PDB ID 2XUE.

Also provided in this disclosure are isolated nucleic acids containing sequences that encode polypeptides described herein. As one example, the nucleic acid can include nucleotides 3797-5422, 3917-4840, or 5003-5254 of SEQ ID NO:2, or sequences at least 85% identical (e.g., 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical) thereto (e.g., over at least 85% of the length of these nucleotides). The sequences at least 85% identical to a sequence disclosed herein encode polypeptides that possess an activity described herein (e.g., JMJD3 catalytic ativity, 2OG binding, Fe(II) binding, or Zn++ binding). The molecular biology methods applied for obtaining such isolated nucleotide fragments are generally known to the person skilled in the art (for standard molecular biology methods see Sambrook et al., Eds., "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), which is incorporated herein by reference). For example, RNA can be isolated from a human cell and cDNA generated applying reverse transcription polymerase chain reaction (RT-PCR) using either random primers (e.g., random hexamers of decamers) or primers specific for the generation of the fragments of interest. The fragments of interest can then be amplified by standard PCR using fragment specific primers. In one embodiment, the isolated polynucleotides are derived from SEQ ID NO: 2. SEQ ID NO:2 encodes the full-length JMJD3 and, thus, polynucleotides coding for the fragments may comprise deletions at the 3'- and/or 5'-ends of the polynucleotide such that the polynucleotide does not encode the full-length protein.

Vectors and Host Cells

A recombinant vector includes any vector known to the skilled person, including plasmid vectors, cosmid vectors, phage vectors such as lambda phage, viral vectors such as adenoviral or baculoviral vectors, or artificial chromosome vectors such as bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), or PI artificial chromosomes (PAC). The vectors include expression as well as cloning vectors. Expression vectors include plasmids as well as viral vectors, and generally contain a desired coding sequence and appropriate DNA sequences necessary for the expression of the operably linked coding sequence in a particular host organism (e.g., bacteria, yeast, plant, insect, or mammal) or in in vitro expression systems. Cloning vectors are generally used to engineer and amplify a certain desired DNA fragment and may lack functional sequences needed for expression of the desired DNA fragments.

In one aspect, the present disclosure relates to a recombinant vector comprising an isolated nucleic acid described herein. Techniques used for the incorporation of nucleic acid sequences of interest into vectors are known (see, e.g., Sambrook et al., 1989, above). Such vectors include any vectors known to the skilled person including plasmid vectors, cosmid vectors, phage vectors such as lambda phage, viral vectors such as adenoviral or baculoviral vectors, or artificial chromosome vectors such as bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), or P1 artificial chromosomes (PAC). The vectors may be expression vectors suitable for prokaryotic or eukaryotic expression. The plasmids may include an origin of replication (ori), a multiple cloning site, and regulatory sequences such as a promoter (constitutive or inducible), transcription initiation site, ribosomal binding site, transcription termination site, polyadenylation signal, and/or selection marker such as antibiotic resistance or auxotrophic marker based on complementation of a mutation or deletion. In one embodiment, the polynucleotide sequence of interest is operably linked to the regulatory sequences. In another embodiment, the vector includes nucleotide sequences encoding epitope-, peptide-, or protein-tags that facilitate purification or detection of the polypeptide of interest. Such tags include, but are not limited to, hemagglutinin- (HA-), FLAG-, c-myc, poly-His (e.g., 6-His (6H)), glutathione-S-transferase- (GST-), maltose-binding-protein- (MBP-), NusA-, and thioredoxin tags, or fluorescent protein-tags such as (enhanced) green fluorescent protein ((E)GFP), (enhanced) yellow fluorescent protein ((E)YFP), red fluorescent protein (RFP) derived from Discosoma species (DsRed) or monomeric (mRFP), cyan fluorescence protein (CFP), and the like. In some embodiments, the tags can be cleaved off the polypeptide, for example, using a protease such as thrombin, Factor Xa, PreScission, TEV protease, and the like. In some embodiments, the tag can be removed with a TEV protease and its recognition site. The recognition sites for such proteases are well known to the person skilled in the art. For example, the seven amino acid consensus sequence of the TEV protease recognition site is Glu-X-X-Tyr-X-Gln-Gly/Ser (SEQ ID NO:3), wherein X may be any amino acid and “Gly/Ser” indicates that the amino acid is either Gly or Ser; for example, the sequence can be Glu-Asn-Leu-Tyr-Phe- Gln-Gly (SEQ ID NO:4). In another embodiment, the vector includes a functional sequence (e.g., a signal sequence) that leads to secretion of the polypeptide into the culture medium of the recombinant host cells or into the periplasmic space of bacteria. The signal sequence usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). In some embodiments, there are processing sites, which can be cleaved either *in vivo* or *in vitro,* encoded between the signal peptide fragment and the foreign gene.

A recombinant host cell refers to a host cell that comprises a polynucleotide/nucleic acid that encodes a polypeptide of interest, i.e., a polypeptide described herein. This nucleic acid may be found inside the host cell (i) freely dispersed as such, (ii) incorporated in a recombinant vector, or (iii) integrated into the host cell genome or mitochondrial DNA. The recombinant cell can be used for expression of a nucleic acid of interest or for amplification of the nucleic acid or the recombinant vector of the disclosure. The recombinant host cell includes the progeny of the original cell into which the

polynucleotide/nucleic acid or the recombinant vector has been introduced (e.g., via transformation, transfection, or infection). A recombinant host cell may be a bacterial cell such as an *E. coli* cell, a yeast cell such as *Saccharomyces cerevisiae* or *Pichia pastoris*, a plant cell, an insect cell such as Sf9 or Hi5 cell, or a mammalian cell. Examples of mammalian cells are Chinese hamster ovary (CHO) cells, green African monkey kidney (COS) cells, human embryonic kidney (HEK293) cells, HELA cells, and the like.

In some embodiments, the disclosure provides a recombinant host cell comprising the isolated nucleic acid or the recombinant vector. The recombinant host cells may be prokaryotic cells such as archea and bacterial cells or eukaryotic cells such as yeast, plant, insect, or mammalian cells. In one embodiment, the host cell is a bacterial cell such as an *E. coli* cell. Methods for introducing an isolated nucleic acid or recombinant vector into a host cell are known in the art. For example, bacterial cells can be readily transformed using, for example, chemical transformation, e.g., the calcium chloride method, or electroporation. Yeast cells may be transformed, for example, using the lithium acetate transformation method or electroporation. Other eukaryotic cells can be transfected, for example, using commercially available liposome-based transfection kits such as LIPOFECTAMINE™ (Invitrogen), commercially available lipid-based transfection kits such as Fugene (Roche Diagnostics), polyethylene glycol-based transfection, calcium phosphate precipitation, gene gun (biolistic), electroporation, or viral infection. In some embodiments, the recombinant host cell expresses the nucleic acid. In another embodiment, expression leads to soluble polypeptide. The polypeptide may be purified using protein purification methods well known to the person skilled in the art, optionally taking advantage of the above-mentioned tags.

Antibodies

Antibody refers to both monoclonal and polyclonal antibodies, i.e., any immunoglobulin protein or portion thereof which is capable of recognizing an antigen or hapten, i.e., a polypeptide possessing JMJD3 catalytic activity (e.g., H3K27 demethylase activity) or the JMJD3 sequence or a peptide thereof, or a polypeptide containing the GATA-like domain described herein or the GATA-like domain sequence or a peptide thereof. Antigen-binding portions of the antibody are also included in the meaning of antibody and may be produced, e.g., by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies.

In one aspect, the present disclosure provides an antibody directed against a polypeptide described herein (e.g., a polypeptide containing a JMJD3 sequence (e.g., with JMJD3 catalytic activity) and/or a GATA-like domain). The antibody may be a monoclonal or polyclonal antibody or a portion(s) (e.g., antigen-binding portion) thereof. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include Fab, Fab', F(ab')2, Fd, Fv, dAb, and complementarity determining region (CDR) fragments (e.g., three CDRs from an antibody heavy chain and/or three CDRs from an antibody light chain (e.g., from the same antibody as the heavy chain)), single-chain antibodies (scFv), chimeric antibodies such

as humanized antibodies, diabodies, and polypeptides that contain at least a portion of an antibody that is sufficient to confer specific antigen binding to the polypeptide. The antibody of the present disclosure can be generated according to standard protocols. For example, a polyclonal antibody may be generated by immunizing an animal such as mouse, rat, rabbit, goat, sheep, pig, cattle, or horse with the antigen of interest optionally in combination with an adjuvant such as Freund's complete or incomplete adjuvant, RIBI (muramyl dipeptides), or ISCOM (immunostimulating complexes) according to standard methods known in the art. Monoclonal antibodies may be generated by methods known in the art, e.g., using hybridoma technology or by screening antibody libraries (e.g., phage or yeast display libraries of antibodies, e.g., human antibodies). Methods of selecting, cloning, and expanding hybridomas are known in the art. The skilled person may refer to standard texts such as "Antibodies: A Laboratory Manual", Harlow and Lane, Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1990), which is incorporated herein by reference, for support regarding generation of antibodies. Phage display is described, for example, in U.S. Pat. No. 5,223,409; Smith (1985) Science 228:1315-1317; WO 92/18619; WO 91/17271; WO 92/20791; WO 92/15679; WO 93/01288; WO 92/01047; WO 92/09690; WO 90/02809; de Haard et al. (1999) J. Biol. Chem 274:18218-30; Hoogenboom et al. (1998) Immunotechnology 4:1-20; Hoogenboom et al. (2000) Immunol Today 2:371-378; and Hoet et al. (2005) Nat Biotechnol. 23(3)344-8. Bacteriophage displaying the protein component can be grown and harvested using standard phage preparatory methods, e.g., PEG precipitation from growth media. After selection of individual display phages, the nucleic acid encoding the selected protein components can be isolated from cells infected with the selected phages or from the phage themselves, after amplification. Individual colonies or plaques can be picked, the nucleic acid isolated and sequenced. Other display formats include cell based display (see, e.g., WO 03/029456), protein-nucleic acid fusions (see, e.g., U.S. Pat. No. 6,207,446), ribosome display (see, e.g., Mattheakis et al. (1994) Proc. Natl. Acad. Sci. USA 91:9022 and Hanes et al. (2000) Nat Biotechnol. 18:1287-92; Hanes et al. (2000) Methods Enzymol. 328:404-30; and Schaffitzel et al. (1999) J Immunol Methods. 231(1-2):119-35), and E. coli periplasmic display (J Immunol Methods. 2006 Jan 20;308(1-2):43-52).

Test Compound

As used herein, the term “test compound” refers to an agent that is being tested for its ability to modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), Fe(II) binding, 2OG binding, or GATA-like domain Zn++ binding . The test compound can be any agent including, but not restricted to, a peptide, a peptoid, a polypeptide (such as an antibody), a lipid, a metal, a nucleotide, a nucleotide analog, a nucleoside, a nucleic acid, an organic compound, an inorganic compound, a small organic or inorganic molecule, a chemical compound (e.g., a low molecular compound, or a high molecular compound), a pharmacological agent, an element, a saccharide, an isotope, a carbohydrate, an imaging agent, a lipoprotein, a glycoprotein, an enzyme, an analytical probe, a polyamine, and combinations and derivatives thereof. A test compound can be naturally occurring or synthetically prepared. It can be isolated from a microorganism, an animal, or a plant, and can be produced recombinantly, or synthesized, e.g., by a chemical method. Examples of

nucleic acid compounds that may be used include RNAi, such as shRNA, siRNA, esiRNA, miRNA, oligo DNA, oligo RNA, a ribozyme, and antisense nucleic acid, such as antisense RNA. The test compound can have a formula weight of less than about 10,000 grams per mole, less than 5,000 grams per mole, less than 1,000 grams per mole, or less than about 500 grams per mole. The test compound of the disclosure may comprise a mixture of substances (e.g., purified or partially purified), for example, the test compound can be an extract (e.g., an extract derived from a marine organism, plant, animal, soil, or a phage display library), or the product of mixed combinatorial syntheses. The mixture can be tested, e.g., in a method described herein, and the component that modulates (e.g., increases or decreases) JMJD3 catalytic activity (e.g., H3K27 demethylase activity), Fe(II) binding, 2OG binding, or GATA-like domain Zn++ binding can be purified from the mixture in a subsequent step.

A test compound can be obtained, for example, using any of the numerous combinatorial library methods known in the art, including but not limited to, biological libraries (such as libraries of peptides or polypeptides (e.g., antibodies), e.g., phage display or eukaryotic cell (e.g., yeast) display libraries) or small molecule libraries, synthetic library methods requiring deconvolution, the “one-bead one-compound” library method, and synthetic library methods using affinity chromatography selection.

Test compounds can be derived or selected from libraries of synthetic or natural compounds. For instance, synthetic compound libraries are commercially available from Maybridge Chemical Co. (Trevillet, Cornwall, UK), ChemBridge Corporation (San Diego, CA, USA), ICCB Known Bioactives library (Enzo Life Sciences, Plymouth Meeting, PA, USA), Aldrich (Milwaukee, WI, USA) ComGenex (Princeton, N.J., USA), Asinex (Moscow, RU), Tripos, Inc. (St. Louis, MO, USA), Otava (Kyiv, UA), Light Biologicals (Shirley, NY, USA), and ChemStar, Ltd. (Moscow, RU). A natural compound library is, for example, available from TimTec LLC (Newark, DE, USA). Libraries of natural compounds in the form of bacterial, fungal, plant and animal cell and tissue extracts can be used. Test compounds can be synthetically produced using combinatorial chemistry either as individual compounds or as mixtures. A collection of compounds made using combinatorial chemistry is a combinatorial library. RNAi (e.g., miRNA, siRNA, esiRNA, or shRNA) libraries can also be prepared, and are available from, for example, Sigma-Aldrich (St. Louis, MO, USA), System Biosciences (Mountain View, CA, USA), Open Biosystems (Huntsville, AL, USA), and Applied Biosystems/Ambion (Austin, TX, USA).

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175; Furka, Int. J. Pept. Prot. Res. 37:487-493 (1991); and Houghton et al., Nature 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication No. WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs et al., Proc. Nat. Acad. Sci. USA 90:6909-6913

(1993)), vinylogous polypeptides (Hagihara et al., J. Amer. Chem. Soc. 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann et al., J Amer. Chem. Soc. 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen et al., J. Amer. Chem. Soc. 116:2661 (1994)), oligocarbamates (Cho et al., Science 261:1303 (1993)), and/or peptidyl phosphonates (Campbell et al., J. Org. Chem. 59:658 (1994)), nucleic acid libraries (see Ausubel, Berger and Sambrook, all supra), peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083), antibody libraries (see, e.g., Vaughn et al., Nature Biotechnology, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al., Science, 274:1520-1522 (1996) and U.S. Pat. No. 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN, January 18, page 33 (1993); isoprenoids, U.S. Pat. No. 5,569,588; thiazolidinones and metathiazanones, U.S. Pat. No. 5,549,974; pyrrolidines, U.S. Pat. Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Pat. No. 5,506,337; benzodiazepines, U.S. Pat. No. 5,288,514, and the like). Additional examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckemmann et al. (1994). J. Med. Chem. 37:2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; and Gallop et al. (1994) J. Med. Chem. 37:1233.

The test compounds of the present disclosure can also be obtained from biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann, R. N. et al. (1994) J. Med. Chem. 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the “one-bead one-compound” library method; and synthetic library methods using affinity chromatography selection. The biological libraries include libraries of nucleic acids and libraries of proteins. Some nucleic acid libraries encode a diverse set of proteins (e.g., natural and artificial proteins; others provide, for example, functional RNA and DNA molecules such as nucleic acid aptamers or ribozymes. A peptoid library can be made to include structures similar to a peptide library. (See also Lam (1997) Anticancer Drug Des. 12:145). A library of proteins may be produced by an expression library or a display library (e.g., a phage or yeast display library).

Libraries of compounds may be presented in solution (e.g., Houghten (1992) Biotechniques 13:412-421), or on beads (Lam (1991) Nature 354:82-84), chips (Fodor (1993) Nature 364:555-556), bacteria (Ladner, U.S. Pat. No. 5,223,409), spores (Ladner U.S. Pat. No. 5,223,409), plasmids (Cull et al. (1992) Proc Natl Acad Sci USA 89:1865-1869) or on phage (Scott and Smith (1990) Science 249:386-390; Devlin (1990) Science 249:404-406; Cwirla et al. (1990) Proc. Natl. Acad. Sci. 87:6378-6382; Felici (1991) J. Mol. Biol. 222:301-310; Ladner, U.S. Pat. No. 5,223,409).

A test compound may optionally comprise a detectable label. Such labels include, but are not limited to, enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds, and bioluminescent compounds.

High throughput screening methods can be used to evaluate test compounds. High throughput screening methods involve providing a library (e.g., a combinatorial chemical, RNAi (e.g., miRNA, siRNA, esiRNA, or shRNA), or peptide library) containing a large number of test compounds (library members). Such libraries are then screened in one or more assays, as described herein, to identify those library members (e.g., particular chemical species or subclasses) that display a desired characteristic activity.

Test compounds and/or libraries thereof can be screened in one or more assays, as described herein, to identify those test compounds and/or library members (particular chemical species or subclasses) that display a desired characteristic activity, e.g., have an effect (e.g., inhibitory effect) on JMJD3 catalytic activity (e.g., H3K27 demethylase activity), Fe(II) binding, 2OG binding, or GATA-like domain Zn++ binding. The compounds thus identified can serve as conventional "lead compounds," e.g., which are varied (e.g., by derivatization), or can themselves be used as modulators of JMJD3 catalytic activity (e.g., H3K27 demethylase activity), Fe(II) binding, 2OG binding, or GATA-like domain Zn++ binding, respectively.

Assays

An assay for identifying and/or selecting test compounds which modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity) can include contacting a solution containing JMJD3 catalytic activity (e.g., containing a polypeptide described herein that contains a JMJD3 catalytic activity) and a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., with at least 3, at least 4, at least 5, at least 6, at least 7, or at least 8 residues on either side of the methylated lysine (e.g., the peptide is methylated, e.g., di- or tri-methylated at the position corresponding to K27)) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) with a test compound and determining the ability of the test compound to modulate (e.g., increase or decrease) JMJD3 catalytic activity (e.g., H3K27 demethylase activity). A co-factor, such as Fe(II) and/or 2OG, can optionally be included. Determining the ability of the test compound to modulate JMJD3 catalytic activity (e.g., H3K27 demethylase activity) can be accomplished, for example, by MALDI-TOF to detect and measure demethylation of the substrate (e.g., conversion of a trimethylated substrate lysine to a dimethylated lysine) by the polypeptide.

JMJD3 binds 2OG and Fe(II). Compounds that interfere with Fe(II) or 2OG binding can decrease (e.g., inhibit) catalytic activity (e.g, H3K27 demethylase activity) and are identifiable by the assays

described herein. Compounds that promote Fe(II) or 2OG binding can increase catalytic activity (e.g, H3K27 demethylase activity) and are identifiable by the assays described herein.

Controls (e.g., negative controls) can also be included in the assay. For example, the amount of demethylation after exposure to a test compound can be compared to the amount of demethylation in the absence of the test compound under identical assay conditions, e.g., identical buffer conditions, reaction time, co-factors, and temperature. For example, a solution containing JMJD3 catalytic activity (e.g., containing a polypeptide described herein that contains a JMJD3 catalytic activity) and a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is di- or tri- methylated at the position corresponding to K27)) can be incubated in the presence and absence of a test compound. After an incubation period, the solutions can be evaluated to determine the amount of demethylation under each condition. If the amount of demethylation in the solution containing the test compound after the period of incubation differs from the amount of demethylation in the solution that does not contain the test compound, the test compound modulates JMJD3 demethylation. If a greater amount of demethylation is measured in the solution containing the test compound as compared to the amount of demethylation in the solution that does not contain the test compound, the test compound increases JMJD3 catalytic (demethylase) activity. If a lesser amount of demethylation is measured in the solution containing the test compound as compared to the amount of demethylation in the solution that does not contain the test compound, the test compound decreases (e.g., inhibits) JMJD3 catalytic (demethylase) activity. The difference in amount of demethylase activity can be a statistically significant difference (e.g., statistical significance by a T test (e.g., student's T test) (e.g., with a probability (p) value of $p<0.05$ or 0.01)), or can be a difference of about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, or about 90%, or greater. As another control, it is possible to compare assay results to a reference, e.g., a reference value, e.g., obtained from the literature, a prior assay, and so forth. Appropriate correlations and art known statistical methods can be used to evaluate an assay result(s).

In some embodiments, the compound that modulates JMJD3 catalytic activity (e.g., H3K27 demethylase activity) decreases (e.g., inhibits) the activity. For example, the compound decreases the JMJD3 catalytic activity (e.g., H3K27 demethylase activity) of the polypeptide by 50%, 60%, 70%, 80%, 90%, or up to 100%, as compared to the activity of the polypeptide without the compound but otherwise with the same reaction conditions, e.g., buffer conditions, reaction time, co-factors, and temperature. In some embodiments, the compound specifically decreases or inhibits the activity of the polypeptide but does not decrease or inhibit the activity of another demethylase, e.g., another Jumonji protein (e.g., JMJD2A, JMJD2B, JMJD2C, JMJD2D, or JMJD6), to the same extent, or not at all.

In another aspect, the disclosure provides a method for identifying a compound which modulates (e.g., increases or decreases) JMJD3 catalytic activity (e.g., H3K27 demethylase activity) of a

polypeptide (that contains a JMJD3 catalytic activity) described herein, the method comprising the steps of

(a) constructing a computer model of the active site defined by the structure coordinates of the polypeptide described herein;

(b) selecting a potential activity modulating compound (e.g., test compound) by a method selected from the group consisting of: (i) assembling molecular fragments into the compound, (ii) selecting a compound (e.g., test compound) from a small molecule database, and (iii) de novo ligand design of the compound;

(c) employing computational means to perform a fitting program operation between computer models of the compound and the polypeptide in order to provide an energy-minimized configuration of the compound with the polypeptide; and

(d) evaluating the results of the fitting operation to quantify the association between the compound and the active site model, thereby evaluating the ability of the compound to associate with the polypeptide; and thereby identifying a compound which modulates JMJD3 catalytic activity.

In some embodiments, the modulating compound decreases (e.g., inhibits) JMJD3 catalytic activity (e.g., H3K27 demethylase activity).

In another aspect, the disclosure provides a method for identifying a compound which modulates (e.g., increases or decreases) Zn++ binding of a polypeptide described herein, the method comprising the steps of

(a) constructing a computer model of the active site defined by the structure coordinates of the polypeptide described herein;

(b) selecting a potential activity modulating compound (e.g., test compound) by a method selected from the group consisting of: (i) assembling molecular fragments into the compound, (ii) selecting a compound (e.g., test compound) from a small molecule database, and (iii) de novo ligand design of the compound;

(c) employing computational means to perform a fitting program operation between computer models of the compound and the polypeptide in order to provide an energy-minimized configuration of the compound with the polypeptide; and

(d) evaluating the results of the fitting operation to quantify the association between the compound and the active site model, thereby evaluating the ability of the compound to associate with the polypeptide; and thereby identifying a compound which modulates Zn++ binding.

In some embodiments, the modulating compound decreases (e.g., inhibits) Zn++ binding.

The co-ordinates disclosed herein may be used for building homology models of variants of JMJD3, orthologues of JMJD3, and homologues of JMJD3, such as UTX and UTY. Such predictive models are valuable, e.g., in light of the higher costs associated with the expression, purification and testing of the many diverse polypeptides that are related to JMJD3 by at least 30% identity. Construction of a

homology model may be achieved through the use of software that is capable of generating three-dimensional co-ordinates of polypeptides or portions thereof from a set of one or more polypeptide structure coordinates. An example for such a computer program is MODELER (Sali and Blundell, 1993, J. Mol. Biol. 234:779-815 as implemented in the Insight II Homology software package (Insight II (97.0), Molecular Simulations Incorporated, San Diego, CA)).

Several methods may be used to screen test compounds for their ability to modulate JMJD3 catalytic activity (e.g., H3K27 demethylase activity) or Zn++ binding. This process may begin by a visual inspection of, for example, a three-dimensional computer model of the polypeptide based on the structural coordinates deposited in, e.g., PDB ID: 2XUE (deposited on October 19, 2010 in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB)). Selected test compounds may then be positioned in a variety of orientations or docked onto the polypeptide (e.g., onto the active site or the GATA-like domain). Docking may be accomplished using software such as Cerius, Quanta, and Sybyl (Tripos Associates, St. Louis, MO), followed by energy minimization and molecular dynamics with standard molecular dynamics force fields such as OPLS-AA, CHARMM, and AMBER. Additional specialized computer programs that may assist the person skilled in the art in the process of selecting suitable compounds include, for example, (i) AUTODOCK (Goodsell et al., 1990, Proteins: Struct., Funct., Genet. 8: 195-202; AUTODOCK is available from The Scripps Research Institute, La Jolla, CA) and (ii) DOCK (Kuntz et al., 1982, J. Mol. Biol. 161 :269-288; DOCK is available from the University of California, San Francisco, CA).

Once suitable compounds have been selected, they can be designed or assembled into a single compound or complex. This manual model building is performed using software such as Quanta or Sybyl. Useful programs aiding the skilled person in connecting individual compounds or fragments include, for example, (i) CAVEAT (Bartlett et al., 1989, in Molecular Recognition in Chemical and Biological Problems, Special Publication, Royal Chem. Soc. 78:182-196; Lauri and Bartlett, 1994, J. Comp. Aid. Mol. Des. 8:51-66; CAVEAT is available from the University of California, Berkley, CA), (ii) 3D Database systems such as ISIS (MDL Information Systems, San Leandro, CA; reviewed in Martin, 1992, J. Med. Chem. 35:2145-2154), and (iii) HOOK (Eisen et al., 1994, Proteins: Struct., Funct., Genet. 19:199-221; HOOK is available from Molecular Simulations Incorporated, San Diego, CA).

Another approach enabled by this disclosure is the computational screening of small molecule databases for compounds that can bind in whole or part to the polypeptide (e.g., the catalytic site or the GATA-like domain). In this screening, the quality of fit of such compounds to the polypeptide (e.g., the active site or the GATA-like domain) may be judged either by shape complementarity or by estimated interaction energy (Meng et al., 1992, J. Comp. Chem. 13:505-524). Alternatively, a potential modulator for the polypeptide, e.g., an inhibitor of a JMJD3 catalytic activity (e.g., H3K27 demethylase activity) or Zn++ binding, may be designed de novo on the basis of the 3D structure of the polypeptide, e.g., deposited in PDB ID: 2XUE (deposited on October 19, 2010 in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB)). There are various de

novo ligand design methods available in the art. Such methods include (i) LUDI (Bohm, 1992, J. Comp. Aid. Mol. Des. 6:61-78; LUDI is available from Molecular Simulations Incorporated, San Diego, CA), (ii) LEGEND (Nishibata and Itai, Tetrahedron 47:8985-8990; LEGEND is available from Molecular Simulations Incorporated, San Diego, CA), (iii) LeapFrog (available from Tripos Associates, St. Louis, MO), (iv) SPROUT (Gillet et al., 1993, J. Comp. Aid. Mol. Des. 7:127-153; SPROUT is available from the University of Leeds, UK), (v) GROUPBUILD (Rotstein and Murcko, 1993, J. Med. Chem. 36:1700-1710), and (vi) GROW (Moon and Howe, 1991, Proteins 11 :314-328).

In addition, several molecular modeling techniques (hereby incorporated by reference) that may support the person skilled in the art in de novo design and modeling of potential modulators and/or inhibitors of polypeptides have been described and include, for example, Cohen et al., 1990, J. Med. Chem. 33:883-894; Navia and Murcko, 1992, Curr. Opin. Struct. Biol. 2:202-210; Balbes et al., 1994, Reviews in Computational Chemistry, Vol. 5, Lipkowitz and Boyd, Eds., VCH, New York, pp. 37-380; Guida, 1994, Curr. Opin. Struct. Biol. 4:777-781. A molecule designed or selected as binding to the polypeptide (e.g., the active site or the GATA-like domain) may be further computationally optimized so that in its bound state it preferably lacks repulsive electrostatic interaction with the region it binds. Such non-complementary (e.g., electrostatic) interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions. Specifically, the sum of all electrostatic interactions between the compound and the region in a bound state preferably make a neutral or favorable contribution to the enthalpy of binding. Specific computer programs that can evaluate a compound deformation energy and electrostatic interaction are available in the art. Examples of suitable programs include (i) Gaussian 92, revision C (Frisch, Gaussian, Incorporated, Pittsburgh, PA), (ii) AMBER, version 4.0 (Kollman, University of California, San Francisco, CA), (iii) QUANTA/CHARMM (Molecular Simulations Incorporated, San Diego, CA), (iv) OPLS-AA (Jorgensen, 1998, Encyclopedia of Computational Chemistry, Schleyer, Ed., Wiley, New York, Vol. 3, pp. 1986-1989), and (v) Insight II/Discover (Biosysm Technologies Incorporated, San Diego, CA). These programs may be implemented, for instance, using a Silicon Graphics workstation, IRIS 4D/35 or IBM RISC/6000 workstation model 550. Other hardware systems and software packages are known to those skilled in the art.

Once a compound has been selected or designed, substitutions may then be made in some of its atoms or side groups in order to improve or modify its binding properties. Generally, initial substitutions are conservative, i.e., the replacement group will approximate the same size, shape, hydrophobicity and charge as the original group. It should be understood that components known in the art to alter conformation should be avoided. Such substituted chemical compounds may then be analyzed for efficiency of fit to the polypeptide by the same computer methods described in detail above.

If computer modeling according to the methods described hereinabove indicates binding of a compound to the polypeptide (e.g., onto the active site or the GATA-like domain), the compound may

be synthesized and optionally the compound or a pharmaceutically acceptable salt thereof may be formulated with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). Thus, the above-described method may comprise the further step of (e) synthesizing the compound and optionally formulating the compound or a pharmaceutically acceptable salt thereof with one or more pharmaceutically acceptable excipient(s) and/or carrier(s). Optionally, the ability of the compound or of a pharmaceutically acceptable salt thereof or of a formulation thereof to modulate, e.g., decrease, JMJD3 catalytic activity (e.g., H3K27 demethylase activity) or Zn++ binding may be tested in vitro or in vivo comprising the further step of (f) contacting the compound with the polypeptide (e.g., and with a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) or the recombinant host cell of the disclosure and to determine (e.g., by MALDI-TOF) the ability of the compound to (i) bind to the polypeptide and/or (ii) to modulate (e.g., decrease, e.g., inhibit) the catalytic activity (H3K27 demethylase activity) or Zn++ binding of the polypeptide. A co-factor, such as Fe(II) and/or 2OG, can optionally be included. The quality of fit of such compounds to the polypeptide may be judged either by shape complementarity or by estimated interaction energy (Meng et al., 1992, J. Comp. Chem. 13:505-524). Methods for synthesizing compounds are well known to the person skilled in the art or compounds may be commercially available. For example, MALDI-TOF, ELISA, Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP) can be used to measure demethylase activity.

In a further aspect, the disclosure provides a method for identifying compounds which modulate (e.g., decrease, e.g., inhibit) JMJD3 catalytic activity (H3K27 demethylase activity), the method comprising (i) contacting a polypeptide described herein (that contains a JMJD3 catalytic activity) (e.g., and with a JMJD3 substrate (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) or a recombinant host cell with a test compound and (ii) analyzing (e.g., by MALDI-TOF) the ability of the test compound to bind to the polypeptide, to modulate (e.g., decrease, e.g., inhibit) JMJD3 catalytic activity (H3K27 demethylase activity) of the polypeptide. A co-factor, such as Fe(II) and/or 2OG, can be included in the assay. For example, the polypeptide and a substrate thereof (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) are contacted in the presence or absence of varying amounts of a test compound and incubated for a certain period of time, for example, for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 60, or 90 minutes. The reaction conditions are chosen such that the polypeptide is catalytically active in the

absence of the test compound. The substrate is then analyzed for demethylation, for example, by MALDI-TOF. ELISA, Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, and chromatin immunoprecipitation (ChIP) can also be used to measure demethylase activity (e.g., assays are available from Active Motif (Carlsbad, CA), and Epigentek (Brooklyn, NY)). Demethylation can also be measured, e.g., using a histone assay (e.g., incubating the polypeptide with core histones, e.g., as described in Miller et al., Genes & Dev. 2008. 22: 2980-2993 or Lee at al., Science 2007: 318: 447 - 450), or by a fluorescent assay. Fluorescent assays to measure H3K27 demethylase activity are available, e.g., from Abnova (Taipei, Taiwan), Active Motif (Carlsbad, CA, USA), and Epigentek (Brooklyn, NY, USA). Antibodies specific to mono-, di- or tri-methylated H3K27 are also available (e.g., from Epigentek (Brooklyn, NY, USA); Abcam (Cambridge, UK)) and can be used to detect and measure levels of mono-, di- or tri-methylated H3K27 as a measure of JMJD3 catalytic activity. Such antibodies can be used, e.g., in ELISA, Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, and chromatin immunoprecipitation (ChIP).

In one embodiment, the interaction between the polypeptide (e.g., and a substrate thereof (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri-methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) and a test compound may be analyzed in form of an enzyme-linked immunosorbent assay (ELISA)-based experiment. The polypeptide may be immobilized on the surface of an ELISA plate and contacted with the test compound. A co-factor, such as Fe(II) and/or 2OG, can be included in the assay. Binding of the test compound may be verified, for example, for proteins, polypeptides, peptides, and epitope-tagged compounds by antibodies specific for the test compound or an epitope tag on the compound. These antibodies might be directly coupled to an enzyme or detected with a secondary antibody coupled to the enzyme that - in combination with the appropriate substrates - carries out chemiluminescent reactions (e.g., horseradish peroxidase) or colorimetric reactions (e.g., alkaline phosphatase). In another embodiment, binding of compounds that cannot be detected by antibodies might be verified by labels directly coupled to the test compounds. Such labels may include enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. In another embodiment, the test compounds might be immobilized on the ELISA plate and contacted with the polypeptide according to the disclosure. Binding of the polypeptide may be verified by a polypeptide-specific antibody (e.g., specific for a mono-, di-, or tri-methylated H3K27) and chemiluminescence or colorimetric reactions as described above. The effects of a test compound on the interaction between the polypeptide and its substrate can be analyzed in like manner. For example, ELISA can be used to detect and/or measure the interaction between the polypeptide and its substrate in the presence and absence of the test compound. The effects of the test compound on the level of interaction can be

analyzed (e.g., determining if the level of interaction decreases or increases in the presence of the test compound).

In a further embodiment, the polypeptide (e.g., and a substrate thereof (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri- methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) may be incubated with a peptide array and binding of the polypeptide to specific peptide spots corresponding to a specific peptide sequence may be analyzed, for example, by polypeptide-specific antibodies, antibodies that are directed against an epitope-tag fused to the polypeptide, or by a fluorescence signal emitted by a fluorescent tag coupled to the polypeptide. A co-factor, such as Fe(II) and/or 2OG, can be included in the assay. The effects of a peptide from the array on the interaction between the polypeptide and its substrate can be analyzed in like manner.

In another embodiment, a recombinant host cell described herein is contacted with a test compound. This may be achieved by co-expression of test compounds (e.g., polypeptides) and verification of interaction, for example, by fluorescence resonance energy transfer (FRET) or co-immunoprecipitation. In another embodiment, directly labeled test compounds may be added to the medium of the recombinant host cells. The potential of the test compound to penetrate membranes and bind to the polypeptide may be, for example, verified by immunoprecipitation of the polypeptide and verification of the presence of the label. In another embodiment, the ability of the test compound to modulate (e.g., decrease, e.g., inhibit) JMJD3 catalytic activity (H3K27 demethylase activity) of the polypeptide is assessed. For example, the purified polypeptide and a substrate thereof (e.g., a peptide corresponding to H3K27 and adjacent sequence (e.g., that is methylated, e.g., di- or tri-methylated at the position corresponding to K27) (e.g., a peptide with the sequence ATKAAR**K**SAPATGGVKKPHRYRPG (SEQ ID NO:5) (the bold underlined K corresponds to K27); or histone 3 (e.g., human histone 3) (e.g., that is methylated, e.g., di- or tri- methylated, at K27)) are contacted in the presence or absence of varying amounts of the test compound and incubated for a certain period of time, for example, for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 60, or 90 minutes. A co-factor, such as Fe(II) and/or 2OG, can be included in the assay. The reaction conditions are chosen such that the polypeptide is catalytically active in the absence of the test compound. The substrate is then analyzed for demethylation, for example, by MALDI-TOF. ELISA, Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, and chromatin immunoprecipitation (ChIP) can also be used to measure demethylase activity (e.g., assays are available from Active Motif (Carlsbad, CA, USA), Epigentek (Brooklyn, NY, USA)). Demethylation can also be measured, e.g., using a histone assay (e.g,, incubating the polypeptide with core histones, e.g., as described in Miller et al., Genes & Dev. 2008. 22: 2980-2993 or Lee at al., Science 2007: 318: 447 - 450), or by a fluorescent assay. Fluorescent assays to measure H3K27 demethylase activity are available, e.g., from Abnova (Taipei, Taiwan), Active Motif (Carlsbad, CA, USA), and Epigentek

(Brooklyn, NY, USA). Antibodies specific to mono-, di- or tri-methylated H3K27 are also available (e.g., from Epigentek (Brooklyn, NY, USA); Abcam (Cambridge, UK)) and can be used to detect and measure levels of mono-, di- or tri-methylated H3K27 as a measure of JMJD3 catalytic activity. Such antibodies can be used, e.g., in ELISA, Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, and chromatin immunoprecipitation (ChIP).

The experiments may be applied in a multi-well plate format and are suitable for high throughput screening of test compounds regarding their ability to modulate (e.g., decrease, e.g., inhibit) JMJD3 catalytic activity (H3K27 demethylase activity). In some embodiments, the above-described methods for identifying compounds which modulate (e.g., decrease, e.g., inhibit) JMJD3 catalytic activity (H3K27 demethylase activity) are performed in a high-throughput setting.

In a further aspect, the disclosure provides a method for identifying compounds which modulate (e.g., decrease, e.g., inhibit) Zn++ binding by the GATA-like domain, the method comprising (i) contacting a polypeptide containing the GATA-liked domain, e.g., a polypeptide described herein (e.g., and Zn++) or a recombinant host cell with a test compound and (ii) analyzing the ability of the test compound to modulate (e.g., decrease, e.g., inhibit) Zn++ binding by the polypeptide. For example, the polypeptide and Zn++ are contacted in the presence or absence of varying amounts of a test compound and incubated for a certain period of time, for example, for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, , 11, 12, 13, 14, 15, 20, 30, 40, 60, or 90 minutes. The reaction conditions are chosen such that the polypeptide binds Zn++ in the absence of the test compound. The reaction is then analyzed for Zn++ binding, e.g., levels of Zn++ (e.g., free Zn++) that is not bound to the GATA-like domain.

Assays to measure Zn++ binding to the GATA-like domain include zinc determination according to Hunt et al, 1985, Anal. Biochem. 146, 150–157. The assay is based on the complex formation of free zinc with the zinc-complexing dye 4-(2-pyridylazo) resorcinol (PAR), which produces an intense red dye (ε500= 66000 M-1 cm-1). The A500 can be monitored using a Beckman DU600 spectrophotometer. Zn++ detecting reagents are also commercially available, e.g., from Invitrogen (Carlsbad, CA, USA).

In another embodiment, the interaction between the GATA-like domain (e.g., a polypeptide containing the GATA-liked domain) (e.g., and Zn++) and a test compound may be analyzed in form of an enzyme-linked immunosorbent assay (ELISA)-based experiment. The polypeptide may be immobilized on the surface of an ELISA plate and contacted with the test compound. Binding of the test compound may be verified, for example, for proteins, polypeptides, peptides, and epitope-tagged compounds by antibodies specific for the test compound or an epitope tag on the compound. These antibodies might be directly coupled to an enzyme or detected with a secondary antibody coupled to the enzyme that - in combination with the appropriate substrates - carries out chemiluminescent reactions (e.g., horseradish peroxidase) or colorimetric reactions (e.g., alkaline phosphatase). In another embodiment, binding of compounds that cannot be detected by antibodies might be verified

by labels directly coupled to the test compounds. Such labels may include enzymatic labels, radioisotope or radioactive compounds or elements, fluorescent compounds or metals, chemiluminescent compounds and bioluminescent compounds. In another embodiment, the test compounds might be immobilized on the ELISA plate and contacted with the polypeptide according to the disclosure. For example, ELISA can be used to detect and/or measure the interaction between the polypeptide and Zn++ in the presence and absence of the test compound. The effects of the test compound on the level of interaction can be analyzed (e.g., determining if the level of interaction decreases or increases in the presence of the test compound).

In a further embodiment, the GATA-like domain (e.g., a polypeptide containing the GATA-liked domain) (e.g., and Zn++) may be incubated with a peptide array and binding of the polypeptide to specific peptide spots corresponding to a specific peptide sequence may be analyzed, for example, by polypeptide-specific antibodies, antibodies that are directed against an epitope-tag fused to the polypeptide, or by a fluorescence signal emitted by a fluorescent tag coupled to the polypeptide. The effects of a peptide from the array on the interaction between the polypeptide and Zn++ can be analyzed in like manner.

In another embodiment, a recombinant host cell described herein is contacted with a test compound. This may be achieved by co-expression of test compounds (e.g., polypeptides) and verification of interaction, for example, by fluorescence resonance energy transfer (FRET) or co-immunoprecipitation. In another embodiment, directly labeled test compounds may be added to the medium of the recombinant host cells. The potential of the test compound to penetrate membranes and bind to the polypeptide may be, for example, verified by immunoprecipitation of the polypeptide and verification of the presence of the label. In another embodiment, the ability of the test compound to modulate (e.g., decrease, e.g., inhibit) Zn++ binding of the polypeptide is assessed. For example, the purified polypeptide and Zn++ are contacted in the presence or absence of varying amounts of the test compound and incubated for a certain period of time, for example, for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 60, or 90 minutes. The reaction conditions are chosen such that the polypeptide can bind Zn++ in the absence of the test compound. The effects of the test compound on Zn++ binding can be analyzed (e.g., determining if the level of binding decreases or increases in the presence of the test compound).

The experiments may be applied in a multi-well plate format and are suitable for high throughput screening of test compounds regarding their ability to modulate (e.g., decrease, e.g., inhibit) Zn++ binding by the GATA-like domain. In some embodiments, the above-described methods for identifying compounds which modulate (e.g., decrease, e.g., inhibit) Zn++ binding by the GATA-like domain are performed in a high-throughput setting.

Crystals

As used herein, the term "crystal" or "crystalline" means a structure (such as a three- dimensional solid aggregate) in which the plane faces intersect at definite angles and in which there is a regular structure (such as internal structure) of the constituent chemical species. The term "crystal" can include any one of: a solid physical crystal form such as an experimentally prepared crystal, a crystal structure derivable from the crystal (including secondary and/or tertiary and/or quaternary structural elements), a 2D and/or 3D model based on the crystal structure, a representation thereof such as a schematic representation thereof or a diagrammatic representation thereof, or a data set thereof for a computer. In one aspect, the crystal is usable in X-ray crystallography techniques. For example, the crystals used can withstand exposure to X-ray beams and are used to produce diffraction pattern data necessary to solve the X-ray crystallographic structure. A crystal may be characterized as being capable of diffracting X-rays in a pattern defined by one of the crystal forms depicted in T. L. Blundell and L. N. Johnson, "Protein Crystallography", Academic Press, New York (1976). The term "unit cell" refers to a basic parallelepiped shaped block. The entire volume of a crystal may be constructed by regular assembly of such blocks. Each unit cell comprises a complete representation of the unit of pattern, the repetition of which builds up the crystal.

The term "space group" refers to the arrangement of symmetry elements of a crystal. In a space group designation, the capital letter indicates the lattice type and the other symbols represent symmetry operations that can be carried out on the contents of the asymmetric unit without changing its appearance.

The term "structure coordinates" refers to a set of values that define the position of one or more amino acid residues with reference to a system of axes. The term refers to a data set that defines the three-dimensional structure of a molecule or molecules (e.g., Cartesian coordinates, temperature factors, and occupancies). S tructural coordinates can be slightly modified and still render nearly identical three-dimensional structures. A measure of a unique set of structural coordinates is the root mean square deviation of the resulting structure. Structural coordinates that render three-dimensional structures (in particular, a three- dimensional structure of an enzymatically active center) that deviate from one another by a root mean square deviation of less than 3 Å, 2 Å, 1.5 Å, 1.0 Å, or 0.5 Å may be viewed by a person of ordinary skill in the art as very similar.

The term "root mean square deviation" means the square root of the arithmetic mean of the squares of the deviations from the mean. It is a way to express the deviation or variation from a trend or object. For purposes of this disclosure, the "root mean square deviation" defines the variation in the backbone of a variant of the polypeptide from the backbone of the polypeptide as defined by the structure coordinates of the polypeptide. As used herein, the term "constructing a computer model" includes the quantitative and qualitative analysis of molecular structure and/or function based on atomic structural information and interaction models. The term "modeling" includes conventional numeric-based molecular dynamic and energy minimization models, interactive computer graphic

models, modified molecular mechanics models, distance geometry, and other structure-based constraint models.

The term "fitting program operation" refers to an operation that utilizes the structure coordinates of a chemical entity, an enzymatically active center, a binding pocket, molecule or molecular complex, or portion thereof, to associate the chemical entity with the enzymatically active center, the binding pocket, molecule or molecular complex, or portion thereof. This may be achieved by positioning, rotating or translating the chemical entity in the enzymatically active center to match the shape and electrostatic complementarity of the enzymatically active center. Covalent interactions, non-covalent interactions such as hydrogen bond, electrostatic, hydrophobic, van der Waals interactions, and non-complementary electrostatic interactions such as repulsive charge-charge, dipole-dipole and charge-dipole interactions may be optimized. Alternatively, one may minimize the deformation energy of binding of the chemical entity to the enzymatically active center.

Crystals can be grown by any method known to the person skilled in the art including, but not limited to, hanging and sitting drop techniques, sandwich-drop, dialysis, and microbatch or microtube batch devices. It would be readily apparent to one of skill in the art to vary the crystallization conditions disclosed above to identify other crystallization conditions that would produce crystals of the polypeptide alone or in complex with a compound. Such variations include, but are not limited to, adjusting pH, protein concentration and/or crystallization temperature, changing the identity or concentration of salt and/or precipitant used, using a different method for crystallization, or introducing additives such as detergents (e.g., TWEEN 20 (monolaurate), LDOA, Brij 30 (4 lauryl ether)), sugars (e.g., glucose, maltose), organic compounds (e.g., dioxane, dimethylformamide), lanthanide ions, or poly-ionic compounds that aid in crystallizations. High throughput crystallization assays may also be used to assist in finding or optimizing the crystallization condition. Microseeding may be used to increase the size and quality of crystals. In brief, micro- crystals are crushed to yield a stock seed solution. The stock seed solution is diluted in series. Using a needle, glass rod or strand of hair, a small sample from each diluted solution is added to a set of equilibrated drops containing a protein concentration equal to or less than a concentration needed to create crystals without the presence of seeds. The aim is to end up with a single seed crystal that will act to nucleate crystal growth in the drop.

Protein crystallization screens can be used, e.g., to identify or select crystallization conditions. Such screens are commercially available, e.g., MORPHEUS™ Screen (Molecular Dimensions, Suffolk, UK); Crystal Screen, Crystal Screen 2, and Crystal Screen HT (Hampton Research, Aliso Viejo, CA, USA); Wizard I, Wizard II, Wizard III, and Wizard IV screens (Emerald BioSystems, Bainbridge Island, WA, USA); and BIOXTAL™ screen kit (XtalQuest, Beijing, CN).

The manner of obtaining structure coordinates, interpretation of the coordinates and their utility in understanding the protein structure, as described herein, are commonly understood by the skilled

person and by reference to standard texts such as J. Drenth, "Principles of protein X-ray crystallography", 2nd Ed., Springer Advanced Texts in Chemistry, New York (1999); and G. E. Schulz and R. H. Schirmer, "Principles of Protein Structure", Springer Verlag, New York (1985). For example, X-ray diffraction data is first acquired, often using cryoprotected (e.g., with 20% to 30% glycerol) crystals frozen to 100 K, e.g., using a beamline at a synchrotron facility or a rotating anode as an X-ray source. Then, the phase problem is solved by a generally known method, e.g., multiwavelength anomalous diffraction (MAD), multiple isomorphous replacement (MIR), single wavelength anomalous diffraction (SAD), or molecular replacement (MR). The substructure may be solved using SFDELXD (Schneider and Sheldrick, 2002, Acta Crystallogr. D. Biol. Crystallogr. (Pt 10 Pt 2), 1772-1779), phases calculated with SHARP (Vonrhein et al., 2006, Methods Mol. Biol. 364:215-30), and improved with solvent flattening and non- crystallographic symmetry averaging, e.g., with RESOLVE (Terwilliger, 2000, Acta Cryst. D. Biol. Crystallogr. 56:965-972). Model autobuilding can be done, e.g., with ARP/wARP (Perrakis et al., 1999, Nat. Struct. Biol. 6:458-63) and refinement with, e.g., REFMAC (Murshudov, 1997, Acta Crystallogr. D. Biol. Crystallogr. 53: 240-255). The skilled person can use the structure coordinates, e.g., PDB ID: 2XUE (deposited on October 19, 2010 in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB)), as input for secondary analysis, including the determination of electrostatic surface potential, which aids in the determination of side groups in test compounds, which are likely to interact with a surface area of the polypeptide of a given electrostatic potential. In order to use the structure coordinates generated for the polypeptide, it is necessary to convert the structure coordinates into a three-dimensional shape. This is achieved through the use of commercially available software that is capable of generating three-dimensional graphical representations of molecules or portions thereof from a set of structure coordinates. An example for such a computer program is MODELER (Sali and Blundell, 1993, J. Mol. Biol. 234:779-815 as implemented in the Insight II Homology software package (Insight II (97.0), Molecular Simulations Incorporated, San Diego, CA)). Such a three-dimensional graphical representation can be used with suitable programs including (i) Gaussian 92, revision C (Frisch, Gaussian, Incorporated, Pittsburgh, PA), (ii) AMBER, version 4.0 (Kollman, University of California, San Francisco, CA), (iii) QUANTA/CHARMM (Molecular Simulations Incorporated, San Diego, CA), (iv) OPLS-AA (Jorgensen, 1998, Encyclopedia of Computational Chemistry, Schleyer, Ed., Wiley, New York, Vol. 3, pp. 1986-1989), and (v) Insight II/Discover (Biosysm Technologies Incorporated, San Diego, CA) to generate graphic representations of, e.g., electrostatic potential. Similarly, the structural information can be combined with information on the conservation of residues at the various amino acid positions to highlight those residues at the surface of the polypeptide.

Incorporation of selenomethionine into proteins in place of methionine aids the structure elucidation of proteins by X-ray crystallography using SAD or MAD. The incorporation of heavy atoms such as selenium helps solving the phase problem in X-ray crystallography.

Examples

Jumonji (Jmj) enzymes are histone demethylases critical for epigenetic mechanisms of cell regulation. A crystal structure of the C-terminus of JMJD3 1141-1682 (del. 1637-1675) reveals that the catalytic domain of JMJD3 is associated with a zinc-binding domain (GATA-like domain) of unusual topology. This provides structural insights into mechanisms of specificity for this hitherto structurally unexplored branch of the Jmj phylogenetic tree important for biological function and drug discovery.

JMJD3 (KDM6B), one of the approximately 30 Jumonji (Jmj) family members found in humans, functions as a specific demethylase of lysine-27 of histone H3 (H3K27) (Hong, S.H. et al. *Proc. Nat. Acad. Sci. USA***104**, 18439-18444 (2007)). Studies have placed JMJD3 at key cell fate decision points in T lymphocytes (Miller, S. A. et al. *Genes Dev.* 22 2980-2993 (2008)) and macrophages. It appears to be involved in acute, externally-driven, inflammatory processes. In macrophages, for example, JMJD3 is rapidly induced through an NF-kB-dependent mechanism in response to bacterial products and inflammatory stimuli (De Santa, F. et al. *Cell* **130**, 1083–1094 (2007)). JMJD3 has also been demonstrated to regulate the differentiation state of the epidermis (Sen, G.L. *Genes Dev.* **22** 1865-1870 (2008)), to activate the tumour suppressor, INK4A-Arf (Sen, G.L. *Genes Dev.* **22** 1865-1870 (2008); Agger, K. et al. *Genes Dev.* **23**, 1171-1176 (2009)) and to be upregulated in prostate cancer (Xiang, Y. et al. *Cell Res.* **17**, 850-857 (2007)).

The Jmj demethylase enzymes have a catalytic (JmjC) domain of the cupin fold that contains a conserved motif of two histidines and either a glutamate or aspartate residue. These key residues coordinate the catalytic iron(II) and co-factor, 2-oxoglutarate (2-OG), required for activity (Cloos, P. et al. *Genes Dev.* **22**, 1115-1140 (2008)). In addition, many Jumonji enzymes possess one or more histone reader domains (e.g., PHD, Tudor, Fbox) that have been shown to be important both for determining substrate specificity and for localising the enzymes to their site of action. The small subfamily of demethylases comprising JMJD3, UTX and UTY have an unique domain architecture amongst the Jmj family and do not contain a recognised reader domain. Much of their N-terminal region consists of sequences of low complexity, amongst which several TPR motifs, implicated in protein-protein interactions, can be detected. At the C-terminus, the catalytic core is followed by a signature for a treble-cleft zinc-finger domain (TCZ) (Cloos, P. et al. *Genes Dev.* **22**, 1115-1140 (2008)) that is not found in other Jumonji enzymes. The function and structure of this domain is unknown.

We have determined the structure and substrate preference of a C-terminal construct of JMJD3 1141-1682 (del. 1637-1675), henceforth referred to as JmjD3t, that contains the JmjC and TCZ domains. This truncate demonstrates activity in vitro and in cells (data not shown and FIG. 6). Substrate profiling of this enzyme alongside JmjD2A (as a 6His-tev-FLAG-JmjD2A 1-350 fragment), with a panel of 198 histone peptides, showed that the preferred substrate for JmjD2A is H3K9Me3, whereas for JmjD3t H3K27Me3 was found to be most rapidly demethylated. In addition, JmjD3t could tolerate

posttranslational modifications on the adjacent residues: dimethylation of R26 and phosphorylation of S28. In cell imaging experiments, JmjD3t was also shown to effectively demethylate di- as well as tri-methylated H3K27, in contrast to a H1390A catalytically deficient mutant (data not shown). This purified crystallographic truncate, therefore, has the same substrate specificity as reported for the full-length enzyme (Burgold, T. *PLoS ONE* 3 e3034 (2008)).

The structure of JmjD3t shows a protein of two domains, the expected catalytic domain and a four helical bundle within which we have observed a Zn-binding region we have termed a GATA-like (GATAL) domain (FIG. 1).

Despite the low sequence homology between JMJD3 and other JmjC domains the spatial distribution of the central beta-strands of its catalytic domain are similar to those observed for JmjD2A, FBX11, and PHF8 (FIG. 3). The surrounding loops and helices, however, differ significantly and prevented the structure solution of JMJD3 by simple molecular replacement methods. The structure was solved by SAD using SelenoMethionine labelled protein, as detailed below.

Within the active site 2-OG makes an extensive network of hydrogen-bonding interactions. The acid of the keto-acid makes direct interactions with the catalytic metal, Ser1398 and Asn1400, while the ketone completes the bi-dentate chelation of the iron. The carboxylate of 2-OG makes four hydrogen bonds to Asn1400, a water, Thr1387 and Lys1381. It is notable that the precise details of the interactions made by this common co-factor differs markedly amongst the Jmjs (FIG. 2) and these differences influence substrate and inhibitor selectivity. For example, pyridine-2,4-dicarboxylic acid (PDCA) is a known inhibitor of JmjD2 family members but has undetectable activity against JMJD3.

Comparison of the 2-OG bound JMJD3 structure with that of PDCA bound to JmjD2A (Protein Data Bank (PDB) ID: 2VD7) suggests possible reasons for the lack of JmjD3 activity. In JmjD2A, Lys206 and Tyr132 hydrogen bond to the 4-carboxylate of PDCA, mimicking the interaction with 2-OG (PDB ID: 2GP5). A direct overlay of these two ligands shows that plane of the two acids differ by approximately 25 degrees (data not shown). A favourable geometry for these hydrogen bonds can be preserved with essentially no visible movement of the residues involved. This is not the case for JMJD3, where the residues that interact with the carboxylate of 2-OG (Thr1387 and Lys1381) originate from quite different regions of the active site. Given the differential disposition of Thr1387 and Lys1381 the 4-carboxylate of PDCA would need to rotate out of plane by more than 45 degrees to maintain these interactions, incurring a significant energetic penalty. Other differences in the active site that could contribute to this differential activity of PDCA are the narrowing of the binding site caused by a Phe185 to Thr1387 substitution between JmjD2A and JMJD3 and the lack of an equivalent residue to Lys241 in JMJD3 to interact with the 2-carboxylate.

The GATAL domain of JMJD3 is an 84 residue insert (1543-1626) between helices 3 and 4 of a C-terminal helical bundle which is also found in PHF8 and FBX11. It is not clear whether this helical

domain is more generally found in JmjCs such as JmjD2A because the crystal structures are often truncated prior to the end of this helical region (FIG. 3). Whilst zinc-binding domains (ZBD) have been seen in the C-terminal region of a number of JmjC domain containing proteins, such as the PHD domain found in PHF8, the architecture of the ZBD in JMJD3 is distinct. The conserved zinc-binding cysteine motif within the GATAL domain is CxxC...CxxC and in JMJD3 consists of Cys1575, Cys1578, Cys1602 and Cys1605. Structurally this domain is related to the erythroid transcription factor, GATA, which also binds a single zinc ion. Therefore we will refer to this as a GATA-like (GATAL) domain. The topology of this fold is illustrated in FIG. 4. The domain consists of a single treble-clef zinc-finger (Grishin, N.V. *Nucleic Acids Res.* **29**, 1703-1714 (2001)) fold elaborated with two extra strands. Treble-clef domains embellished with additional structural elements have also been observed in the ubiquitin-binding domain ZnF UBP (Seigneurin-Berny, D. et al. *Mol. Cell. Biol.* **21**, 8035–8044 (2001)), although the additional strands in ZnF UBP zinc fingers have a different topology to that found in JMJD3.

GATA-type zinc-fingers bind DNA and have also been observed to be involved in protein-protein interactions (Gamsjaeger, R., et al. *Trends in Biochemical Sciences* 32, 63-7014 (2007)). Therefore we speculate that this GATAL domain may have a role in specificity, by directly interacting with the substrate and/or localising JMJD3 to its chromatin target complex.

Comparison of the GATAL domain with that from GATA-1 (Omichinski, J.G., *et al. Biochemistry* **31**, 3907-3917 (1992)) shows that the long C-terminal tail of GATA-1 that binds to the DNA minor groove is absent in JMJD3 (FIG. 7). Instead, the region C-terminal to the helix of the treble clef forms one of the additional strands of the GATAL domain. This suggests that a DNA-binding function for this domain is unlikely. GATA domains are also known to be mediators of protein-protein interactions (Gamsjaeger, R. et al. *Trends in Biochemical Sciences* **32**, 63-70 (2007)) as exemplified by the interaction between GATA-1 and Friend of GATA (FOG) (Chu Kong Liew et al. *Proc. Nat. Acad. Sci. USA* **102**, 583-588 (2005)). The GATAL domain of JMJD3 may therefore be an additional recognition site for the histone substrate or a docking site of a yet unidentified partner protein. Inspection of the surface of JMJD3 shows that there is a groove in the surface of the zinc-binding domain that is aligned with the active site. The ability of a histone peptide to simultaneously bind at the catalytic site and at the GATAL interface, where a peptide corresponding to H3(27-34) has been modelled: A histone peptide comprising residues ^{27}KAARKSAP (SEQ ID NO:7) of H3 was modelled into the cleft observed in the helical and GATAL domain to illustrate the distance and length of peptide that may be recognised (data not shown). Few structures of GATA protein complexes have been elucidated. An overlay of the GATA-1 from the GATA-1/FOG (PDB ID: 1Y0J) complex with the JMJD3 GATAL domain (FIG. 5) demonstrates that this would be able to participate in a similar protein-protein complex, as the corresponding interaction surface is exposed and can accommodate the FOG peptide.

A GATAL domain is predicted to be present in the structures of the homologous proteins UTX and UTY in which the zinc chelating cysteines are completely conserved. A number of residues that appear critical for the structural integrity of this domain and its interface with the catalytic core are also fully conserved. Interestingly, there is significant variability of the residues that line the GATAL groove. Given that JMJD3, UTY and UTX are all H3K27 demethylases, it is intriguing to speculate that this variation may contribute to specificity beyond the active site and the ability of these proteins to carry out distinct functions.

In conclusion, we have solved the first structure of a JMJD3 subfamily of H3K27 demethylases. The distinct active site geometry of JMJD3 and the presence of an integral zinc-binding domain provides novel insights into specificity determinants within and outside the catalytic domain with implications for its biological function and drug discovery.

Cloning and Expression of JMJD3-6H 1141-1682 (del 1637-1675)

JMJD3 DNA encoding the catalytic domain (residues 1141–1682) and lacking residues 1637-1675 (a splice variant deletion; GenBank Accession No. BC009994) was cloned into the pFB-HTb vector (Invitrogen) using BamHI and XhoI restriction sites, which encodes for a protein with an N-terminal 6His (6H) tag followed by a TEV-protease cleavage site. pFB1-HTb-JMJD3-6H 1141-1682 (del 1637-1675) was transposed into the baculovirus genome using the BAC-to-BAC technology (Invitrogen). Bacmid DNA was transfected into *Spodoptera frugiperda* (Sf9) cells using Cellfectin II (Invitrogen), and expression was performed at a 1L scale in Excel 420 media (SAFC Biosciences). The culture, at a cell concentration of $3.8x10^6$ cells/ml, was infected with P1 recombinant Baculovirus at a nominal multiplicity of infection of 3 and incubated for 48 hours. The cells were removed from the media by centrifugation at 2500g for 20 minutes, and the cell pellet was frozen for subsequent purification.

Expression of JMJD3-6H 1141-1682 (1637-1675 deletion) with Seleno-Methionine Incorporation Using Sf9 Cells.

The Baculovirus Expression Vector System (BEVS) was used to produce recombinant protein with Seleno-Methionine incorporated to enable phasing of the x-ray diffractions. *Spodoptera frugiperda* (Sf9) cells were cultivated in suspension using Excell 420 media (SAFC Biosciences) on a shaker incubator (Kuhner) at 27^0C, 80 RPM. In mid exponential growth, the cells were removed from the growth media by centrifugation at 250g for 10 minutes on a bench top centrifuge (Heraeus). The cells were inoculated into the expression media, “Methionine Free” Excell 420 (SAFC Biosciences), at a cell concentration of $1.5x10^6$ cells/ml. The culture was incubated for 27 hours using the described conditions to metabolise the residual amounts of methionine present from hydrolysates in serum free insect media. The culture, at a cell concentration of $3.8x10^6$ cells/ml, was infected with P2 recombinant Baculovirus at a nominal multiplicity of infection of 3 and incubated for 19 hours. 25mg/ml Seleno-Methionine (Sigma-Aldrich) was added to the culture to a concentration of 50mg/litre and the culture incubated for a further 26 hours. The cells were removed from the media by

centrifugation at 2500g for 20 minutes on aforementioned bench top centrifuge. The cell pellet was frozen for subsequent purification.

Purification of JMJD3-6H 1141-1682 (del 1637-1675)

The pellet from the Baculovirus culture was resuspended in buffer A (20mM Tris pH 8.5, 300mM NaCl, 10% glycerol, 1ul/ml Protease Inhibitor Cocktail Set III (Calbiochem 539134) and 10uM Fe(NH4)2(SO4)2).

Cells were lysed by Dounce Homogenisation, on ice, and centrifuged at 100,000 x g for 90 minutes at 4°C. The 100,000 x g supernatant was applied to a HisTRAP HP Column (GE Healthcare 17-5248-02). The column was washed with ten column volumes of buffer A, followed by ten column volumes of buffer A containing 20mM Imidazole. Bound protein was eluted from the column using a linear gradient of 20-250mM Imidazole over twenty column volumes. The JMJD3 protein was eluted between 100mM and 200mM Imidazole.

Eluted JMJD3 protein from the HisTRAP column was concentrated fourfold (Amicon Ultrafree-15 30kDa, Millipore UFC903024) and loaded onto a HiLoad 26/20 Superdex 200 prep grade size exclusion column (GE Healthcare 17-1069-01), equilibrated with buffer B (20mM Tris pH 8.0, 150mM NaCl, 5% glycerol, 0.5mM TCEP, 2mM α-Ketoglutaric acid sodium salt (Sigma K1875) and 10uM Fe(NH4)2(SO4)2). Fractions containing JMJD3 were pooled and concentrated (Ultrafree-15 30kDa).

Protein identity was confirmed by peptide mass fingerprinting and predicted molecular weight confirmed by mass spectrometry.

Purification of SelMet JMJD3-6H 1141-1682 (del 1637-1675)

The pellet from the Baculovirus culture was resuspended buffer A (20mM Tris pH 8.5, 300mM NaCl, 10% glycerol, 1ul/ml Protease Inhibitor Cocktail Set III (Calbiochem 539134) and 10uM Fe(NH4)2(SO4)2).

Cells were lysed by Dounce Homogenisation, on ice, and centrifuged at 100,000 x g for 90 minutes at 4°C. The 100,000 x g supernatant was applied to a HisTRAP HP Column (GE Healthcare 17-5248-02). The column was washed with ten column volumes of buffer A, followed by ten column volumes of buffer A containing 20mM Imidazole. Bound protein was eluted from the column using a linear gradient of 20-250mM Imidazole over twenty column volumes. The JMJD3 protein was eluted between 60mM and 150mM Imidazole.

Eluted JMJD3 protein from the HisTRAP column was diluted 20-fold with buffer C (20mM Tris pH 8.5, 10% glycerol and 10uM Fe (NH4)2(SO4)2) and applied to a HiTRAP Q HP column (GE Healthcare 17-1154-01). The column was washed with ten columns of buffer C and eluted using a linear 0-1M

NaCl gradient over twenty column volumes. The JMJD3 protein was eluted between 350mM and 450mM NaCl.

JMJD3 protein eluted from the HiTRAP Q column was concentrated threefold (Amicon Ultrafree-15 30kDa, Millipore UFC903024) and loaded onto a HiLoad 26/20 Superdex 200 prep grade size exclusion column (GE Healthcare 17-1069-01), equilibrated with buffer B (20mM Tris pH 8.0, 150mM NaCl, 5% glycerol, 0.5mM TCEP, 2mM α-Ketoglutaric acid sodium salt (Sigma K1875) and 10uM Fe(NH4)2(SO4)2). Fractions containing JMJD3 were pooled and concentrated (Ultrafree-15 30kDa).

Protein identity was confirmed by peptide mass fingerprinting.

Cloning and Expression of JMJD3-6H 1141-1682

JMJD3 1141-1682 with no internal deletion was expressed as a 6H-Tev-Flag construct. JMJD3 DNA encoding the catalytic domain (residues 1141–1682) was amplified by PCR to contain a Flag tag at the N-terminus and the PCR product cloned into the pFB-HTb vector (Invitrogen) using BamHI and XhoI restriction sites, which encodes for a protein with an N-terminal 6His tag followed by a TEV-protease cleavage site then a Flag tag. pFB1-HTb-Flag-JMJD3 1141-1682 was transposed into the baculovirus genome using the BAC-to-BAC technology (Invitrogen). Bacmid DNA was transfected into Spodoptera frugiperda (Sf9) cells using Cellfectin II (Invitrogen), and expression was performed at a 1L scale in Excel 420 media (SAFC Biosciences). The culture, at a cell concentration of $3.8x10^6$ cells/ml, was infected with P1 recombinant Baculovirus at a nominal multiplicity of infection of 3 and incubated for 48 hours. The cells were removed from the media by centrifugation at 2500g for 20 minutes, and the cell pellet was frozen for subsequent purification.

Purification of JMJD3-6H 1141-1682

Purification in the presence of Fe2+ ions: The pellet from the Baculovirus culture was resuspended in buffer A (20mM HEPES pH 8.0, 300mM NaCl, 10% glycerol, 1ul/ml Protease Inhibitor Cocktail Set III (Calbiochem 539134) and 10uM Fe(NH4)2(SO4)2).

Cells were lysed by Dounce Homogenisation, on ice, and centrifuged at 100,000 x g for 90 minutes at 4°C. The 100,000 x g supernatant was applied to a HisTRAP HP Column (GE Healthcare 17-5248-02). The column was washed with ten column volumes of buffer A, followed by ten column volumes of buffer A containing 20mM Imidazole. Bound protein was eluted from the column using a linear gradient of 20-250mM Imidazole over twenty column volumes. The JMJD3 protein was eluted between 130mM and 200mM Imidazole.

Eluted JMJD3 protein from the HisTRAP column was concentrated (Amicon Ultrafree-15 30kDa, Millipore UFC903024) and half was loaded onto a HiLoad 26/20 Superdex 200 prep grade size exclusion column (GE Healthcare 17-1069-01), equilibrated with buffer B (20mM HEPES pH 8.0,

150mM NaCl, 5% glycerol, 1mM DTT and 10uM Fe(NH4)2(SO4)2). The protein precipitated on the column.

The remaining protein was loaded onto a HiLoad 26/20 Superdex 200 prep grade size exclusion column (GE Healthcare 17-1069-01), equilibrated with buffer C (20mM Tris pH 8.0, 150mM NaCl, 5% glycerol and 1mM DTT). The protein precipitated on the column.

Purification in the absence of Fe2+ ions: The pellet from the Baculovirus culture was resuspended in buffer A (20mM Tris pH 8.0, 300mM NaCl, 10% glycerol and 1ul/ml Protease Inhibitor Cocktail Set III (Calbiochem 539134)).

Cells were lysed by Dounce Homogenisation, on ice, and centrifuged at 100,000 x g for 90 minutes at 4°C. The 100,000 x g supernatant was applied to a HisTRAP HP Column (GE Healthcare 17-5248-02). The column was washed with ten column volumes of buffer A, followed by ten column volumes of buffer A containing 20mM Imidazole. Bound protein was eluted from the column using a linear gradient of 20-250mM Imidazole over twenty column volumes. The JMJD3 protein was eluted between 140mM and 205mM Imidazole.

Protein identity was confirmed by peptide mass fingerprinting and predicted molecular weight confirmed by mass spectrometry. (Observed the following MWs: 65581 Da (-8 Da), could be minus N-terminal Methionine, +1PO4 and +1 acetylation; 65694 Da (+105 Da), could be minus N-terminal Methionine and +3PO4; 65808 Da (+219 Da), could be +1 acetylation and +2PO4 and 65888 Da (+299 Da), could be + 1 acetylation and +3PO4).

Eluted JMJD3 protein from the HisTRAP column was concentrated threefold (Amicon Ultrafree-15 30kDa, Millipore UFC903024) and half was loaded onto a HiLoad 26/20 Superdex 200 prep grade size exclusion column (GE Healthcare 17-1069-01), equilibrated with buffer D (20mM Tris pH 8.0, 150mM NaCl, 5% glycerol, 1mM DTT and 1ul/ml Protease Inhibitor Cocktail Set III (Calbiochem 539134)). The protein precipitated on the column.

Co-crystallisation of Native and Selenomethionine Labelled JMJD3 6His 1141-1682 (del 1637-1675)

Protein was supplied post SEC (size exclusion chromatography)/concentration at 10mg/mL in 20mM Tris pH 8.0, 150mM NaCl, 5% glycerol, 0.5mM TCEP, 10uM Fe(NH4)2(SO4)2 and 2mM 2-OG. Protein was mixed with an equal volume of Morpheus screen (Molecular Dimensions) condition #A12 (0.03M MgCl2, 0.03M CaCl2, 0.1M tris HCl/bicine pH 8.5, 12.5%v/v MPD, 12.5%w/v PEG1K, 12.5%w/v Peg3350) to a total drop volume of 300nL in a sitting drop MRC plate. Crystals grew at 20°C over a period of one week to an approximate size of 120x70x10 microns and were harvested into mother liquor plus 15% glycerol before flash freezing.

SelMet Data Collection

X-ray diffraction data were collected at the European Synchrotron Radiation Facility at beam line ID14-4. Datasets were collected at 100 K using an ADSC Q315 CCD detector with an X-ray beam transmission of 50% and an oscillation range of 1°. A fluorescence energy scan was taken in the vicinity of the selenium edge. The crystal used for data collection was sufficiently long that a strategy was adopted to try to ensure a highly redundant peak dataset could be collected for use in a SAD experiment, as well as further datasets for use in a MAD experiment. Datasets were collected at four different wavelengths using a different part of the crystal for each wavelength. Firstly, 500° of data were collected at the inflection wavelength (0.3 s exposure), then 360° at a high energy remote wavelength (0.3 s exposure) and then 220° at a low energy remote wavelength (0.4 s exposure) to provide a high resolution dataset. Finally 2000° of data were collected at the peak wavelength (0.2 s exposure). The data were processed and scaled using the HKL2000 suite of programs (HKL Research, Inc., Charlottesville, VA, USA). Structure factors were derived from the reflection intensities using the CCP4 suite of programs (Collaborative Computational Project No. 4). The crystal space group is P1 with unit cell dimensions a = 61.2 Å, b = 65.2 Å, c = 77.5 Å, α = 86.1°, β = 67.2°, γ = 68.3°. Data collection statistics are given in **Table A** for the two datasets that were used in the crystal structure determination.

The structure was solved using the Single-wavelength Anomalous Dispersion (SAD) technique. The program AUTOSHARP (Vonrhein, Clemens; Blanc, Eric; Roversi, Pietro; Bricogne, Gerard. Automated structure solution with autoSHARP. Methods in Molecular Biology (Totowa, NJ, United States) (2007), 364(Macromolecular Crystallography Protocols, Volume 2) was used for heavy atom detection, heavy atom refinement, phasing and density modification, in conjunction with associated CCP4 programs. The expected number of 8 selenium sites per molecule with two molecules in the asymmetric unit was used for initial heavy atom detection. The resulting solvent flattened 2.5 Å experimental electron density map showed clear protein features such as α-helices and β-strands. The CCP4 program CAD (CCP4 Acta Cryst. D50, 760-763, 1994) was used to combine the high resolution dataset with the AUTOSHARP phase information. RESOLVE (P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, Acta Cryst. D66, 213-221 (2010) was then used to perform further density modification using the high resolution data, making use of the two-fold non-crystallographic symmetry present, and then to perform automatic model building. The final 2.0 Å experimental electron density map from RESOLVE was excellent. The autobuilt coordinates provided a good starting point for manual rebuilding of the structure using the model building program Coot Emsley P, Cowtan K Acta Crystallogr. D60, 2126-2132, 2004). Structure refinement was carried out using AUTOBUSTER (Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Smart OS, Vonrhein C and Womack, TO (2009) BUSTER, version 2.8.0. Cambridge, United Kingdom: Global Phasing Ltd.), utilising TLS refinement and automatic water fitting. The initial refinement statistics were R = 0.324, Rfree = 0.311. The presence of 2-oxo-glutarate was clearly seen in the

active site. Addition of this co-factor and water followed by iterative rounds of modeling using Coot (Crystallographic Object-Oriented Toolkit) and refinement (refmac/buster) gave a final model with statistics given in **Table B** (R=0.174, Rfree=0.214).

Table A - Data collection statistics

Parameter[a]	Peak dataset	High resolution dataset
Wavelength (Å)	0.9795	0.9810
Data scaled (°)	1,080	220
Resolution (Å)	40-2.49 (2.53-2.49)	40-2.00 (2.03-2.00)
Observations	394,980	143,151
Unique reflections	35,072	61,840
Redundancy	11.3 (8.5)	2.3 (2.1)
Completeness (%)	97.5 (90.4)	89.9 (56.1)
Average $I/\sigma I$	34.9 (6.2)	12.3 (1.5)
R_{merge}[b]	0.077 (0.278)	0.065 (0.507)

[a] Data for the highest resolution shell are given in parentheses.

[b] $R_{merge} = \Sigma|I_j - \langle I_j \rangle|/\Sigma\langle I_j \rangle$.

Table B – Xray Refinement Statistics

Parameter[a]	SeMet high resolution dataset
Data collection	
Space Group	P1
Cell Dimensions	
a,b,c (Å)	61.21, 65.15 , 77.46
α, β, γ (°)	86.09, 67.19, 68.26
Resolution (Å)	40-2.00 (2.03-2.00)
R_{merge}[b]	0.065 (0.507)
Average $I/\sigma I$	12.3 (1.5)
Completeness (%)	96.2
Redundancy	2.3 (2.1)
Refinement	
Resolution (Å)	20.15-2.00
No. Reflections	143,151
No. Unique Reflections	61,840 (1961)
R_{work}/R_{free}	0.17/0.21
No. atoms	7439
Protein	6841
Ligand/ion	20/14

Water	564
B-factors	
Protein	43.77
Ligand/ion	47.03/102.87
Water	48.21
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.089

[a] Data for the highest resolution shell are given in parentheses.

[b] $R_{merge} = \Sigma|I_j - \langle I_j \rangle|/\Sigma\langle I_j \rangle$.

JMJD3 6His 1141-1682 (del 1637-1675) Crystal Structure

The crystal structure coordinates for this polypeptide containing JMJD3 catalytic activity was deposited on October 19, 2010 in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) and assigned PDB ID 2XUE.

The structure of JMJD3t (JMJD3-6H 1141-1682 (del 1637-1675)) shows a protein of two domains, the expected catalytic domain and a four helical bundle within which we have observed a Zn-binding region we have termed a GATA-like (GATAL) domain. Within the active site the catalytic metal is chelated by H1390, E1392, and H1470. The structure of the 2-oxoglutarate (2-OG) co-factor crystallised within JMJD3 reveals that 2-OG makes an extensive network of hydrogen-bonding interactions. The acid of the keto-acid makes direct interactions with the catalytic metal, Ser1398 and Asn1400, while the ketone completes the bi-dentate chelation of the iron. The carboxylate of 2-OG makes four hydrogen bonds to Asn1400, a water, Thr1387, and Lys1381. It is notable that the precise details of the interactions made by this common cofactor differ markedly amongst the Jmjs and these differences are likely to influence substrate and inhibitor selectivity.

Additional Polypeptides

Additional polypeptides that were prepared and crystallized include:

6H-JMJD3 1142-1682 (del 1638-1675)

JMJD3 1141-1682 (del 1637-1675) (PreScission protease cleaved protein to remove C-terminal 6H tag)

JMJD3 1141-1682 (del 1637-1675) (TEV cleaved protein to remove C-terminal 6H tag)

6H-tev-FLAG-JMJD3 1141-1682 (del 1637-1675)

JMJD3-6His 1065-1682 (del 1637-1675) (crystals too small to be confirmed as protein)

Treble Clef Zinc-Binding Fold: GATA-Like Domain in JMJD3

All sequence numbers are based on the numbering of residues taken from NCBI reference sequence entries NM_001080424 (nucleotide sequence) and NP_001073893 (amino acid sequence). The domain is found at residues 1543-1626 of the sequence shown in NP_001073893 and is encoded by nucleotides 5003-5254 of the sequence shown in NM_001080424.

The presence of a treble-clef zinc-binding domain (Grishin, N.V. (2001) Nucleic Acids Research 29, 1703-1714) following the catalytic domain of JmjD3 was suggested (Cloos, P. et al. Genes & Development 22, 1115-1140 (2008)). However the precise structure and possible function of this domain was unknown. The structure of this GATA-like zinc-binding fold has not been previously determined in other JmJc domain-containing proteins of known structure.

The domain is inserted into a bundle of four helices which are seen in a number of JMJc domain structures, for example FBXL11 (PDB accession code 2YU1). The domain is formed by extension of one of the helices in the bundle, the zinc binding treble clef fold and some additional beta strands. The topology of the additional beta strands is not the same as seen in the ZnF-UBP domain type. The extent of the domain is defined from a superposition of the structure of JMJD3 with the structure of FBXL11 (PDB accession code 2YU1). The first residue of the domain is taken as the first residue within the extended helix that is not superimposed with a helical residue in 2YU1, that is serine 1543 in JMJD3. The last residue of the domain is defined as the residue before the first match of residue position and conformation in the final helix of the four helix bundle, glutamate 1626 in JMJD3.

This domain may have a role in specificity, either by directly interacting with the substrate or localising JMJD3 to its chromatin target complex. The domain is proximal to the active site in the catalytic domain, and could interact with the H3 histone tail and/or JMJD3 partner proteins or DNA.

***In vitro* Substrate Peptide Identification for JMJD2a and JMJD3**

Assay Principle: A formaldehyde dehydrogenase (FDH) coupled continuous fluorescent assay was used to measure the rate of demethylation of the Alta BioScience H3 and H3, H4 N terminal peptide libraries by JMJD2a and JMJD3. The rate of demethylation was monitored by measuring the rate of oxidation of formaldehyde to formic acid by FDH, which reduces APAD+ (a more stable analog of NAD+) to APADH, which has an excitation wavelength of 360 nm and an emission wavelength of 485 nm.

Materials: All reagents and solvents used were of the highest quality commercially available and were purchased from Sigma-Aldrich (Poole, UK) unless otherwise stated. The H3 and H3, H4 N terminal libraries were purchased from Alta BioScience, University of Birmingham, UK. For JMJD3, a 6His-tev-FLAG-JMJD3 1141-1682 (del 1637-1675) was used and prepared as described above. For JMJD2A, a 6His-tev-FLAG-JMJD2A 1-350 fragment was used.

Method: The Alta Bioscience H3 and H3, H4 N terminal libraries were initially made up to 1 mM in DMSO. 1 µL of each of the peptides were placed into 384, black, low volume assay plates (Greiner Bio-One, UK). 5 µL of a substrate/coupling solution (50 µM αketoglutaric acid disodium salt hydrate, 100 µM Fe(NH2)2(SO4)2, 100 µM (+)-Sodium L-ascorbate, 2 mM 3-acetylpyridine adenine dinucleotide (APAD+) and 8 U/mL FDH) in assay buffer (50 mM Hepes, pH 7 and 150 mM NaCl) was then added. The reaction was initiated with the addition of 5 µL of the relevant enzyme (2 µM JMJD2A or 5 µM JMJD3) in the assay buffer above. Fluorescence intensity measurements were taken at 20 seconds intervals for 10 minutes, using a Tecan Saffire spectrofluorimeter (ex.λ_{360nm}, em.λ_{485nm}, gain = 70, temperature at 37°C).

The results demonstrate that the 6His-tev-FLAG-JMJD3 1141-1682 (del 1637-1675) polypeptide is able to demethylate H3K27, and thus has the same substrate specificity as the full length JMJD3 protein. This JMJD3 polypeptide has preference for di- and tri-methylated H3K27. 6His-tev-FLAG-JMJD3 1141-1682 (del 1637-1675) demethylated H3K27Me3 in the absence or other posttranslational modification and also when adjacent residues R26 and S 28 were methylated and phosplorylated, respectively. The 6His-tev-FLAG-JMJD2A 1-350 polypeptide has a preference for tri-methylated H3K9. (Data not shown).

Methods for Cell Culture and Transduction for Cellular Imaging Assay

Human embryonic kidney (HEK) 293 cells were maintained at 37ºC in DMEM/F-12 medium (Invitrogen) supplemented with 10% fetal calf serum (FCS). HEK cells were cultured to 80% confluency prior to being transduced in suspension at a multiplicity of infection of 100 with either FLAG-tagged JMJD3 1141-1682 wild type or FLAG-tagged JMJD3 1141-1682 demethylase dead (H1390A) BacMam viruses. Cells for immuno-fluorescence and cellular imaging were seeded onto cover slips and Greiner poly-D-lysine coated 384 well plates respectively. 24 hours post-transduction, the respective cell lines expressing the ectopic JMJD3 wild type and JMJD3 H1390A mutant forms were fixed with 4% paraformaldehyde for 15 minutes. After washing with phosphate buffered saline supplemented with 1% Tween-20 (PBS-T), cells were blocked for 30 minutes at room temperature in the presence of PBS-T containing 1% bovine serum albumin. The samples were stained with anti-FLAG monoclonal antibody (Sigma M2 F135), as well as one of the rabbit polyclonal histone modification antibodies corresponding to anti-H3K27$(Me)_3$ (Millipore 07-449), H3K27$(Me)_2$ (Abcam Ab24684), or H3K27$(Me)_1$ (Millipore 07-448). Cells were washed with PBS-T prior to counter-staining with Hoechst dye, Alexa FLUOR® 488 Goat anti-mouse IgG and Alexa FLUOR® 594 Donkey anti-rabbit IgG to visualise the nuclear compartment, JMJD3 expressing cells and respective histone modification marks, respectively. The immuno-fluorescence pictures were captured on a standard fluorescence microscope. Quantification of the histone modification marks was achieved by imaging the cells on the Cellomics Array scan and establishing the nuclear Alexa FLUOR® 594 fluorescence intensity in a sub-population of cells expressing either the FLAG-tagged JMJD3 1141-1682 wild type or FLAG-tagged JMJD3 1141-1682 demethylase dead (H1390A) mutant BacMam viruses. The results are shown in FIG. 6.

Results: FLAG-tagged JMJD3 1141-1682 wild type demethylates H3K27$(Me)_3$ and H3K27$(Me)_2$. Quantification of the respective H3K27 methyl states on an Arrayscan demonstrates FLAG-tagged JMJD3 1141-1682 substrate preference for the tri- and di- methyl valencies.

JMJD3 MALDI-TOF Demethylase Assay Protocol

The JMJD3 MALDI-TOF mass spectrometric assay monitors demethylation of a histone H3 peptide containing tri-methylated K27 by recombinant Jumonji D3 (JMJD3) demethylase enzyme. The effects of a test compound in JMJD3 catalytic activity can be evaluated in this assay.

JMJD3 Protein Preparation

JMJD3-6H 1141-1682 (del 1637-1675) protein was prepared as detailed above.

H3K27Me_3 Peptide Synthesis

Peptide sequence: ATKAAR**K**SAPATGGVKKPHRYRPG (the bold underlined lysine corresponds to K27 in human histone 3 and is tri-methylated) (SEQ ID NO:5): ATKAAR-K(TriMe)-SAPATGGVKKPHRYRPG (SEQ ID NO:6)

The peptide was supplied by Cambridge Research Biochemicals. In detail, the protected peptide was assembled on a solid-phase synthesiser using preloaded Wang resin and utilising standard Fmoc synthesis protocols. The crude peptide was obtained after cleavage from the resin with a mixture of trifluoroacetic acid (TFA), triisopropylsilane and water (95:2.5:2.5) for 3 hours at room temperature and was then purified using a C18 reverse-phase column utilising a 0.1%TFA-buffered water/acetonitrile gradient. The resulting fractions were analysed and fractions which were >95% pure by analytical HPLC and giving the correct molecular weight (mw) (by MALDiTOF mass spectroscopy) were pooled and freeze dried. The final material was analysed by HPLC and MALDiTOF mass spectroscopy.

The trimethyl lysine was incorporated into the sequence during assembly as Fmoc-Lys(TriMe)-OH. This was made by reacting a suitably protected alpha-nitrogen lysine with an excess of methyl chloride or bromide to form the quaternary salt.

Compound plate preparation

Compounds were diluted to 10mM in DMSO, and a 1:3, 11 point serial dilution performed across a 384 well hibase plate (Greiner Bio-one, Stonehouse, UK). 100nl of this dilution series was then transferred into a 384 well V-base plate (Greiner Bio-one) using the ECHO™ acoustic dispenser (Labcyte Inc, Sunnyvale, CA, USA) to create the assay plate. 100nl DMSO was added to column 6 to generate a high control.

Assay method

10ul of a 1% trifluoroacetic acid (TFA) solution was added to column 18 of compound plates to generate a 100% inhibition control, before 5ul of an enzyme solution containing 1.2uM 6His-Tev-Flag-JMJD3t and 200uM CHAPS (Sigma-Aldrich, St. Louis, MO) diluted in a buffer of 20mM Tris HCL and 150mM NaCl (both Sigma-Aldrich) was added to each well of the plate using a Multidrop COMBI® dispenser (Thermo Fisher Scientific, Waltham, MA, USA). This gives a final assay concentration of 600nM 6His-Tev-Flag-JMJD3t enzyme and 100uM CHAPS.

Plates were incubated for 15 minutes at room temperature, after which time 5ul of a substrate solution containing 4mM ascorbic acid (Sigma-Aldrich), 100uM ammonium iron (II) sulphate (Fisher Scientific Waltham, MA, USA), 70uM alphaketoglutarate (Sigma-Aldrich) and 30uM H3K27Me_3 peptide, all diluted in a buffer of 50mM Hepes, was added to each well of the plate using a Multidrop COMBI® dispenser, to initiate the reaction. Final assay concentrations are 2mM ascorbic acid, 50uM ammonium iron (II) sulphate, 35uM alphaketoglutarate and 15uM H3K27Me_3 peptide.

After precisely 8 minutes, the reaction was quenched by the addition of 10ul of a 1% TFA solution to all wells except column 18 (to which this was previously added) using a Multidrop COMBI®dispenser.

Samples were diluted 1:100 in 5% acetonitrile using the BIOMEK® FX (Beckman Coulter, Brea, CA, USA) and 1ul of this dilution transferred on to a 384 well PAC target (Bruker Daltonics, Bremen, Germany) again using the BIOMEK® FX. The target was washed by immersing in a solution of 10mM ammonium phosphate (Sigma-Aldrich) and 0.1% TFA for 10 seconds before being dried at room temperature and analysed by MALDI-TOF (Ultraflex III, Bruker Daltonics) to generate relative peak areas of H3K27Me_3 substrate and H3K27Me_2 product.

Claims

1. A method for identifying a compound which modulates JMJD3 catalytic activity, the method comprising:

(a) constructing a computer model of a polypeptide comprising a JMJD3 sequence defined by the structure coordinates of the polypeptide, wherein the polypeptide possesses a catalytic activity of JMJD3; and wherein the JMJD3 sequence consists of:

(i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1;

(ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1;

(iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1; or

(iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1;

(b) selecting a test compound by a method selected from the group consisting of:

(i) assembling molecular fragments into the compound, (ii) selecting a compound from a small molecule database, and (iii) de novo ligand design of the compound;

(c) employing computational methods to perform a fitting program operation between computer models of the test compound and the polypeptide in order to provide an energy-minimized configuration of the compound with the polypeptide; and

(d) evaluating the results of the fitting operation to quantify the association between the test compound and the polypeptide computer model, thereby evaluating the ability of the test compound to associate with the polypeptide.

2. The method of claim 1, wherein the structure coordinates are deposited in Protein Data Bank (PDB) ID: 2XUE.

3. The method of claim 1, wherein the JMJD3 sequence consists of (i) a sequence at least 85% identical to residues 1141-1682 of SEQ ID NO:1.

4. The method of claim 1, wherein the JMJD3 sequence consists of (ii) a sequence at least 85% identical to residues 1141-1682 with a deletion of residues 1637-1675 (del 1637-1675) of SEQ ID NO:1.

5. The method of claim 1, wherein the JMJD3 sequence consists of (iii) a sequence at least 85% identical to a sequence that initiates at any of residues 253 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1.

6. The method of claim 1, wherein the JMJD3 sequence consists of (iv) a sequence at least 85% identical to a sequence that initiates at any of residues 247 to 1140 of SEQ ID NO:1 and terminates at residue 1682 of SEQ ID NO:1, wherein the sequence comprises a deletion of residues 1637-1675 of SEQ ID NO:1.

7. The method of claim 1, wherein the polypeptide comprises a heterologous sequence.

8. The method of claim 1, wherein the method further comprises contacting the test compound with the polypeptide or a recombinant host cell comprising the polypeptide; and evaluating if the test compound modulates JMJD3 catalytic activity of the polypeptide.

9. The method of claim 1, wherein a decrease in JMJD3 catalytic activity indicates that the test compound decreases JMJD3 catalytic activity.

10 The method of claim 1, wherein an increase in the JMJD3 catalytic activity indicates that the test compound increases JMJD3 catalytic activity.

11. The method of claim 1, wherein the JMJD3 catalytic activity is determined by measuring demethylase activity.

12. The method of claim 1, wherein the demethylase activity is determined by MALDI-TOF, enzyme-linked immunosorbent assay (ELISA), Western blotting, immunofluorescence, immunohistochemistry, immunoprecipitation, or chromatin immunoprecipitation (ChIP).

13. The method of claim 1, wherein the test compound is comprised in a compound library.

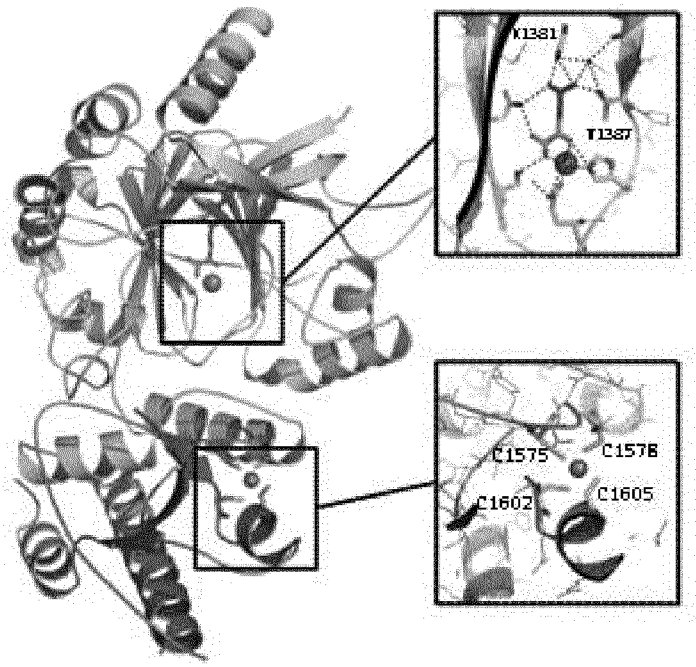


FIG. 1

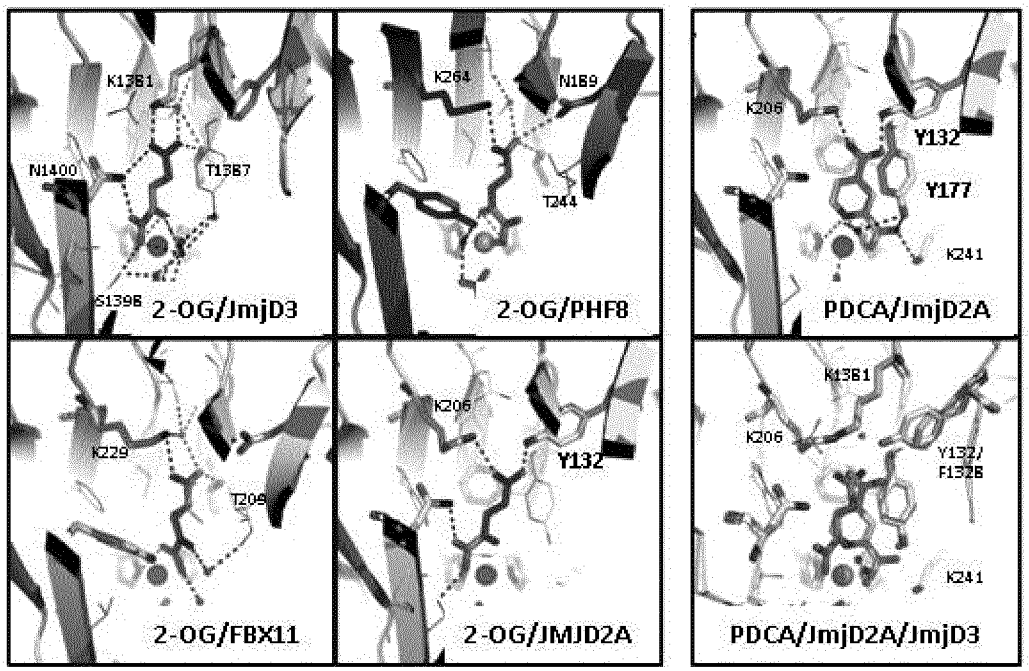

K1381
N1400
T1387
S1398
2-OG/JmjD3
K264
N189
T244
2-OG/PHF8
K206
Y132
Y177
K241
PDCA/JmjD2A
K229
T209
2-OG/FBX11
K206
Y132
2-OG/JMJD2A
K1381
K206
Y132/
F132B
K241
PDCA/JmjD2A/JmjD3

FIG. 2

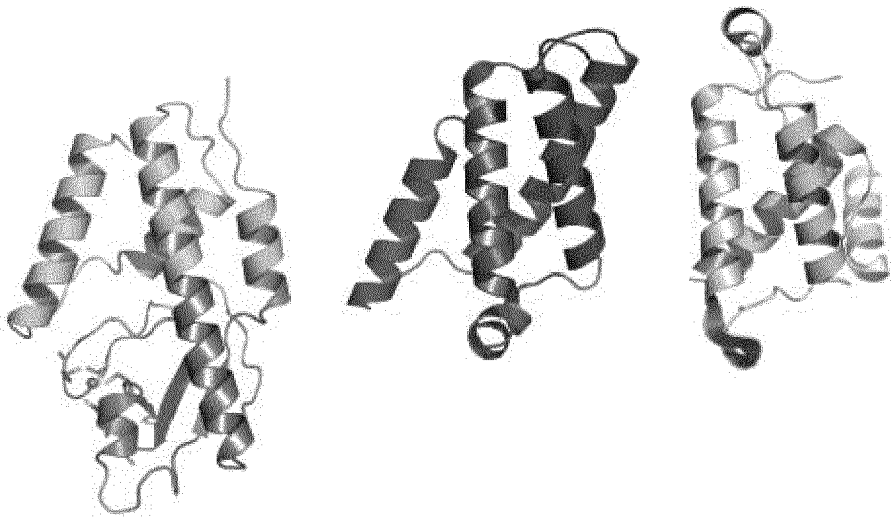


FIG. 3

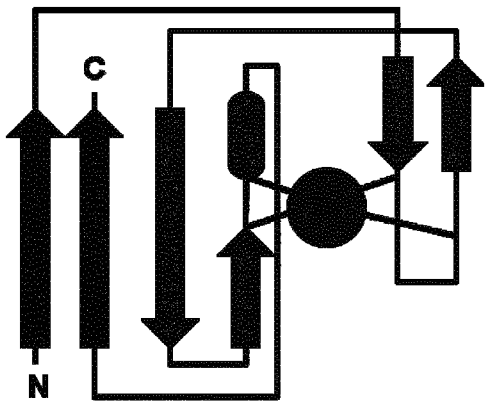


FIG. 4

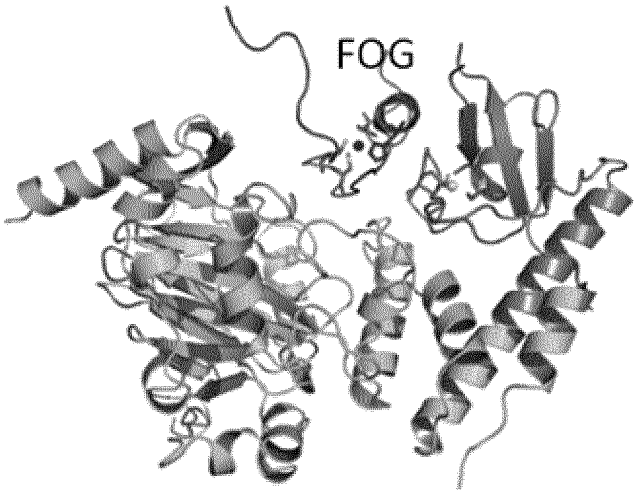


FIG. 5

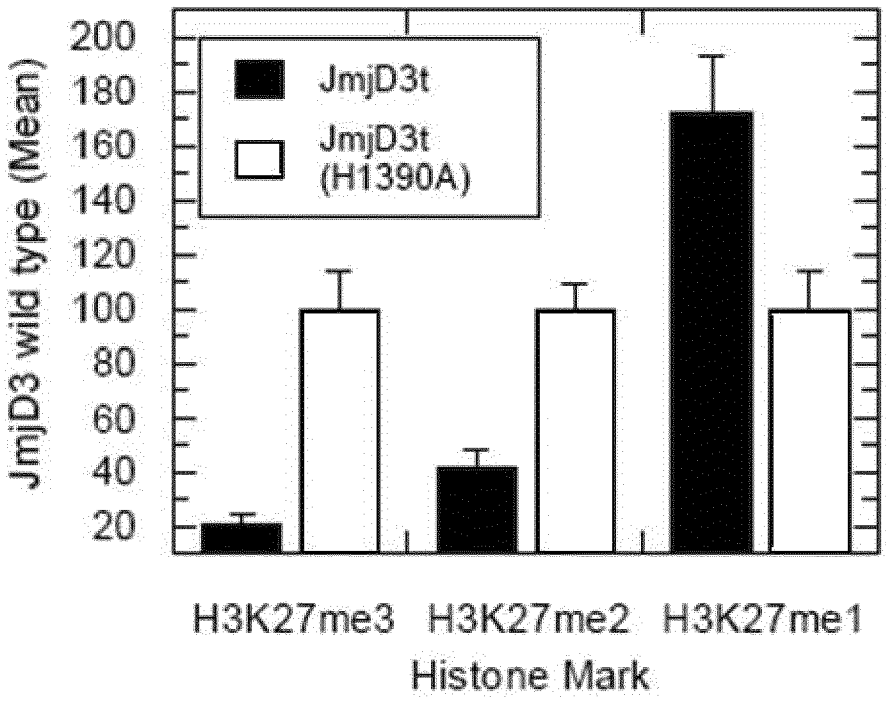
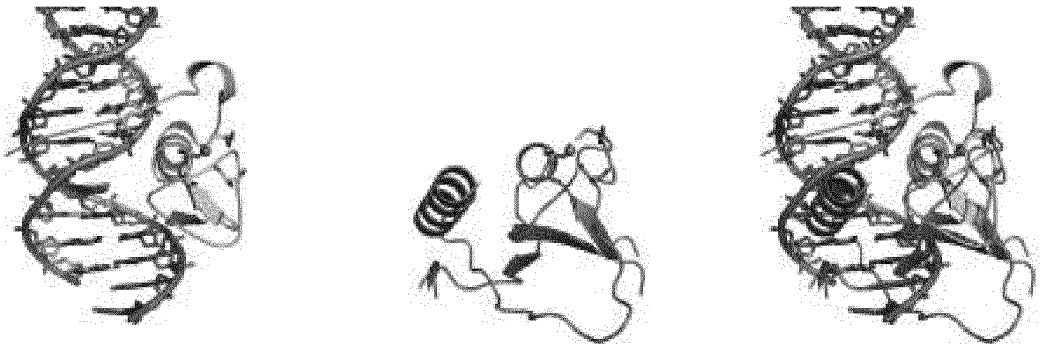

JmjD3 wild type (Mean)
200
180
160
140
120
100
80
60
40
20
JmjD3t
JmjD3t (H1390A)
H3K27me3
H3K27me2
H3K27me1
Histone Mark

FIG. 6



(i). GATA-1 with DNA

(ii). GATAL domain of JmjD3 in the orientation as GATA-1 in i.

(iii). Overlay of GATAL with GATA-1

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/068087

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/16
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	Che, K.H. et al: "Crystal structure of the human JMjd3 Jumonji domain", pdb , 12 November 2010 (2010-11-12), XP002670337, Retrieved from the Internet: URL:http://www.rcsb.org/pdb/explore.do?structureId=2xxz [retrieved on 2012-02-23] the whole document ----- -/--	1-13

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 February 2012	12/03/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Lüdemann, Susanna

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/068087

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	SENGOKU TORU ET AL: "Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A", GENES & DEVELOPMENT, vol. 25, no. 21, 1 November 2011 (2011-11-01), pages 2266-2277, XP009156824, the whole document	1-13
A	KARL AGGER ET AL: "UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development", NATURE, vol. 449, no. 7163, 11 October 2007 (2007-10-11), pages 731-734, XP55020283, ISSN: 0028-0836, DOI: 10.1038/nature06145 the whole document	1-13
A	P. A.C. CLOOS ET AL: "Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease", GENES & DEVELOPMENT, vol. 22, no. 9, 1 May 2008 (2008-05-01), pages 1115-1140, XP55020284, ISSN: 0890-9369, DOI: 10.1101/gad.1652908 the whole document	1-13