STRUCTURAL BASIS OF LSD1-CoREST SELECTIVITY IN HISTONE H3 RECOGNITION

Federico Forneris1*, Claudia Binda1*, Antonio Adamo2, Elena Battaglioli2 and Andrea Mattevi1

1 Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy;
2 Dipartimento di Biologia e Genetica per le Scienze Mediche, Università di Milano, via Viotti 3/5, 20133 Milano, Italy;

Correspondence to: Andrea Mattevi, Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy; Tel. +39-0382-985534, Fax: +39-0382-528496, E-mail: mattevi@ipv.gen.unipv.it

Running title: LSD1–histone peptide complex

Coordinates and Structure Factors have been deposited with the Protein Data Bank with the accession code 2V1D.

* These authors equally contributed to this work.

Histone demethylase LSD1 regulates transcription by demethylating Lys4 of histone H3. The crystal structure of the enzyme in complex with CoREST and a substrate-like peptide inhibitor highlights an intricate network of interactions and a folded conformation of the bound peptide. The core of the peptide structure is formed by Arg2, Gln5 and Ser10 which are engaged in specific intramolecular H-bonds. Several charged side chains on the surface of the substrate-binding pocket establish electrostatic interactions with the peptide. The three-dimensional structure predicts that methylated Lys4 binds in a solvent inaccessible position in front of the flavin cofactor. This geometry is fully consistent with the demethylation reaction being catalyzed through a flavin-mediated oxidation of the substrate amino-methyl group. These features dictate the exquisite substrate specificity of LSD1 and provide a structural framework to explain the fine tuning of its catalytic activity and the active role of CoREST in substrate recognition.

Lysine methylation is among the most well characterized histone modifications and its existence has been known since the early days of chromatin research (1,2). This type of epigenetic mark provides a huge potential for functional responses in that it can occur in different forms (mono-, di-, and tri-methylation) and on different histone sites each having a specific physiological meaning. Histone methylation has been long thought to be a low-turnover epigenetic mark but the recent discovery of histone demethylases (3,4) challenged this view demonstrating that histone lysine methylation can be actively and dynamically regulated. Two classes of histone demethylases have been uncovered; the enzymes of the JmjC family use iron as cofactor whereas lysine-specific-demethylase 1 (LSD1) employs FAD as prosthetic group (5).

LSD1 catalyses the oxidative demethylation of mono- and di-methyl Lys4 of histone H3, generating hydrogen peroxide and formaldehyde (3,4). The enzyme is implicated as a key component of distinct co-activator and co-repressor complexes in a surprisingly wide range of cellular processes where it participates in the dynamic transition of transcriptional programs (6). Its catalytic activity is finely tuned by the epigenetic marks present on the H3 N-terminal tail (7) and by other protein partners, such as CoREST, that form stable a complex with the enzyme (8,9). The three dimensional structure of LSD1 in its native state (10,11) and in complex with the LSD1-binding domain of CoREST (12) revealed that the catalytic centre is located in the core of the enzyme main body. A
protruding tower domain consisting of two remarkably long helices forms the docking site for the co-repressor protein (Fig. 1a).

Here, we describe the structural analysis of LSD1-CoREST bound to a 21 amino acid H3 peptide in which pLys4 ("p" is for peptide) is mutated to Met. The structural analysis illuminates the molecular properties that enable LSD1 to function as a key transcriptional regulator through both its histone demethylase activity and its highly specific binding to the H3 N-terminal tail.

**EXPERIMENTAL PROCEDURES**

**Protein purification**

*E. coli* cells over-producing GST-fused LSD1-binding domain of CoREST (residues 305-482) were harvested 8 h after addition of 0.25 mM isopropyl β-D-thiogalactopyranoside. Cells (10 g) were disrupted with an Emulsiflex C3 cell disruptor and the extract loaded onto a Glutathione Sepharose 4B column (GE Healthcare) followed by extensive washing with 50 mM Tris/HCl pH 7.2, 300 mM NaCl, 1 mM DTT, 1 mM EDTA. GST-tag cleavage was performed by incubating the column-bound protein with 80 units of Prescision Protease (GE Healthcare) overnight at 4° C. Next, the eluted protein was gel-filtered using a Superdex 200 column equilibrated in 5% (w/v) glycerol, 25 mM potassium phosphate pH 7.2. Human recombinant LSD1 was expressed and purified following published protocols (7). The protein lacks the N-terminal 122 amino acids (with reference to the sequence deposited in NCBI database under the accession code NP_055828). Purified CoREST was mixed with LSD1 in a 1.5:1 molar ratio for 1 h and gel-filtered on a Superdex 200 column equilibrated in 5% (w/v) glycerol, 25 mM potassium phosphate pH 7.2. The elution peak had the retention volume expected for the CoREST-LSD1 complex whose formation was further verified by SDS-PAGE analysis.

**Enzymatic assays**

Enzymatic activities were measured under aerobic conditions by using a peroxidase-coupled assay on a Cary 100 UV/Vis spectrophotometer (7). Peptides were purchased from Thermo Electron Corporation. Their purity was greater than 90% as checked by analytical high-pressure liquid chromatography and mass spectrometry. Inhibition was tested in the presence of varied concentrations (2-100 μM) of mono-methylated H3-K4 substrate and pLys4Met peptide (Table 1).

**Crystallization and structure determination**

The LSD1-CoREST complex was crystallized by the hanging-drop vapor diffusion method at 20° C by mixing equal volumes of 100 μM LSD1-CoREST solution in 5% (w/v) glycerol, 25 mM potassium phosphate pH 7.2 with reservoir solutions containing 1.2-1.3 M sodium/potassium tartrate, 100 mM N-(2-acetamido)-2-iminodiacetic acid pH 6.5. Crystals are isomorphous to those obtained by Yang et al. (12); their cell parameters and space group are reported in Table 2. Crystals were soaked in a solution consisting of 1.6 M sodium/potassium tartrate, 100 mM N-(2-acetamido)-2-iminodiacetic acid pH 6.5, 10% glycerol, and 2 mM pLys4Met peptide (Fig. 1b) for 3 h and then flash-cooled in a stream of gaseous nitrogen at 100 K. X-ray diffraction data were collected at the X06SA beam-line of the Swiss Light Source (Villigen, Switzerland). Data processing and scaling (Table 2) were carried out using MOSFLM (13) and programs of the CCP4 package (14). The structure of the LSD1-CoREST complex (PDB entry 2IW5) (12) after removal of all water atoms was used as initial model for refinement. Unbiased 2Fo-Fc and Fo-Fc maps were used to manually build the protein-bound peptide inhibitor (Figure 1b) whose conformation was clearly defined by the electron density (Figure 1b). Crystallographic refinements were performed with Refmac5 (15) and manual re-building was done with Coot (16). Refinement statistics are listed in Table 2. Pictures were produced with PyMol (www.pymol.org), Bobscript (17), and Raster3d (18).

**RESULTS AND DISCUSSION**

Our structural analysis is based on human LSD1-CoREST crystals that are isomorphous to those reported by Yang et al. (12) although they were grown under slightly different conditions (see Experimental Procedures). Structure of the ternary complex was obtained using a peptide bearing Met in place of Lys in position 4. With respect to the wild-type peptide ($K_i = 1.8 \mu M$) (7), such a mutation leads to a 30-fold increase in binding affinity.


\( K_i = 0.05 \ \text{µM} \) making the pLys4Met substrate-analogue a strong inhibitor and an ideal candidate for structural studies (Table 1). The electron density map allowed us to unambiguously position the 16 N-terminal amino acids of the ligand (Fig. 1b). Peptide binding does not induce any conformational change as indicated by root-mean-square deviation of 0.4 Å from ligand-free LSD1-CoREST (799 equivalent Ca atoms). The peptide binds in a funnel-shaped pocket (Figs. 1a and 2a) adopting a folded conformation in which three structural elements can be identified; a helical turn (residues 1-5) located in front of the flavin, a sharp bent (residues 6-9), and a more extended stretch (residues 10-16) that remains partly solvent exposed on the rim of the binding pocket (Fig. 2a). All amino acids fall in the allowed regions of the Ramachandran plot and, therefore, do not appear to adopt strained conformations. The core of the peptide structure is formed by pArg2, pGln5 and pSer10 which are engaged in several intramolecular interactions (Figs. 2b and 2c). The side chain of pArg2 is crucial in that it establishes H-bonds with the carbonyl oxygens of pGly12 and pGly13 and the side chain of pSer10. Moreover, all peptide residues define an intricate network of very specific intermolecular contacts with LSD1 often involving electrostatic interactions (Fig. 2). Among these, the N-terminal amino group of pAla1 favorably interacts with the C-terminus of helix 524-540 (Figs. 1a and 2c) whereas the pArg2 and pArg8 form salt bridges with the carboxylate groups of Asp556 and Glu379, respectively. The side chains of pLys9 and pLys14 are partly disordered pointing towards the solvent (Fig. 2a). However, they both can establish favorable electrostatic interactions with a cluster of negatively charged residues on LSD1 surface that involve Asp553, Asp556, Asp557, and Glu559 (Fig. 2a).

The three-dimensional structure of the ternary complex explains several biological properties of LSD1. The residue in position 4 binds right in front of the cofactor with the side chain pointing towards the flavin ring of FAD (Fig. 2b). This residue is made solvent inaccessible by the peptide C-terminal portion that effectively seals the active center. A modeling analysis (Fig. 2d) in which pMet4 is replaced by di-methylLys shows that the di-methyl group can be easily accommodated in the space between the peptide and the flavin. The predicted position of the pLys4–N-CH$_3$ group (the site of oxidative demethylation) is at about 3 Å distance from the reactive N5 atom of the flavin and the geometry of the flavin-methylamino group interactions are identical to those of flavo-dehydrogenases/oxidases (19). In particular, the methyl-amino group of the substrate is fixed in position through the interaction with an aromatic side chain (Tyr761; Figs. 2b and 2d) as typically found in other flavin-dependent amine oxidases. Thus, the structure is fully compatible with the proposed oxidative mechanism for demethylation that involves flavin-mediated oxidation of the N-CH$_3$ group of methylLys4.

An astonishing feature is the intricacy of the interactions between LSD1 and the histone peptide, which is consistent with the enzyme exquisite specificity (Fig. 2). LSD1 is not active towards peptides shorter than 16 amino acids and optimal activity is detected with peptides comprising the N-terminal 21 amino acids of H3 (7). Consistently, the structure shows that the first 16 amino acids of the H3 peptide are all specifically interacting with the protein. After pPro16, the electron density is too poorly defined to allow model building (Fig. 1b). We speculate that flexibility of amino acids 17-21 may be required to allow the tail protruding from the nucleosomal particle to reach the binding pocket. The complexity of the peptide-enzyme interactions rationalizes the ability of LSD1 to recognize epigenetic modifications on H3. pSer10 is solvent inaccessible being part of the intramolecular interactions that stabilize the peptide conformation. Phosphorylation of this residue is predicted to disrupt this binding geometry in keeping with the fact that pSer10-phosphorylated peptides are unable to bind to LSD1 (7). Likewise, both pArg2 and pArg8 side chains establish several inter- and intramolecular H-bonds explaining the reduced binding affinity caused by arginine methylation (20). pLys9 and pLys14 are solvent accessible and close to a cluster of negatively charged residues on LSD1 surface (Fig. 2a). Indeed, methylation of pLys9 (that preserves the side chain positive charge) does not affect the normal kinetic parameters and, therefore, it is not expected to perturb the peptide binding (7). Conversely, lysine acetylation with consequent removal of the positive charge is predicted to diminish the favorable interaction with the surrounding negatively charged protein environment.
accounting for the 6-fold increase in $K_m$ caused by pLys9 acetylation and the lack of LSD1 activity on iperacetylated peptides and nucleosomes (9,20). Taken together, these structural data explain the unique combination of specificity and affinity of LSD1 for deacetylated/dephosphorylated H3 that enables the enzyme to demethylate Lys4 and to have non-enzymatic functions in the recruitment of other chromatin modifying proteins and protein complexes (6,21,22).

The co-repressor protein CoREST is necessary to make LSD1 able to act on intact nucleosomal particles (8). Moreover, CoREST-bound LSD1 exhibits a two-fold increase in the rate of catalysis measured with peptide substrate (Table 1). The three-dimensional structure offers a clue to understand these functional properties indicating that the co-repressor indirectly takes part in substrate binding (Figs. 2a and 2c). In particular, residues 14-16 of the histone peptide are in contact with helix 372-395 (Figs. 1a and 2a) of LSD1 which in turn extensively interacts with CoREST. The shortest distance between a CoREST residue and the peptide is 7.2 Å (Pro311-pGly13). These structural features combined with the functional data suggest that CoREST may have a dual role in optimizing LSD1 activity: it contributes to the overall architecture of the peptide-binding site and it is very likely to physically interact with the nucleosomal particles facilitating their recognition by LSD1.

Acknowledgements
The financial support by the Italian Ministry of Science (PRIN06 and FIRB programs) and the Italian Association for Cancer Research (AIRC) is gratefully acknowledged. C. Binda is supported by a Young Investigator Fellowship from “Collegio Ghislieri”, Pavia. We thank Dr. C. Schulze-Briese of SLS for excellent support in data collection.

REFERENCES
1. Sims, R.J. 3rd, Nishioka, K., and Reinberg, D. (2003) Trends Genet. 19, 629-639
2. Kouzarides, T. (2007) Cell 128, 693-705
3. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004) Cell 119, 941–953
4. Forneris, F., Binda, C., Vanoni, M.A., Mattevi, A., and Battaglioli, E. (2005) FEBS Lett. 579, 2203 – 2207
5. Klose, R.J., and Zhang, Y. (2007) Nat. Review Mol. Biol. 8, 307-318
6. Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F.L., Korach, K.S., Glass, C.K., Fu, X.D., and Rosenfeld, M.G. (2007) Nature 446, 882-887
7. Forneris, F., Binda, C., Vanoni, M.A., Battaglioli, E., and Mattevi, A. (2005) J. Biol. Chem. 280, 41360-41365
8. Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005) Nature 437, 432-435
9. Shi, Y., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005) Mol. Cell 19, 1-8
10. Stavropoulos, P., Blobel, G., and Hoelz, A. (2006) Nat. Struct. Mol. Biol. 13, 626-632
11. Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y., and Lei, M. (2006) Proc. Natl. Acad. Sci. U S A 103, 13956-13961
12. Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006) Mol. Cell. 23, 377-387
13. Leslie, A. G. W. (1999) Acta Crystallogr. D55, 1696-1702
14. Collaborative Computational Project, Number 4. (1994) Acta Crystallogr. D50, 760-767
15. Murshudov, G. N., Vagin, A.A., and Dodson, E.J. (1997) Acta Crystallogr. D53, 240-255
16. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D60, 2126-2132
17. Esnouf, R.M. (1999) Acta Crystallogr. D55, 938-940
18. Merritt, E.A., and Bacon, D.J. (1997) Methods in Enzymology 277, 505-524
19. Fraaije, M.W., and Mattevi, A. (2000) Trends in Biochemical Sciences 25, 126-132
20. Forneris, F., Binda, C., Dall’Aglio, A., Fraaije, M.W., Battaglioli, E., and Mattevi, A. (2006) J. Biol. Chem. 281, 35289-35295
21. Wissmann, M., Yin, N., Muller, J.M., Greschik, H., Fodor, B.D., Jenuwein, T., Vogler, C., Schneider, R., Gunther, T., Buettner, R., Metzger, E., and Schule, R. (2007) Nat. Cell. Biol. 9, 347-353
22. Ballas, N., Battaglioli, E., Atouf, F., Andres, M.E., Chenoweth, J., Anderson, M.E., Burger, C., Moniwa, M., Davie, J.R., Bowers, W.J., Federoff, H.J., Rose, D.W., Rosenfeld, M.G., Brehm, P., and Mandel, G. (2001) Neuron 31, 353-365

FIGURE LEGENDS

Figure 1. Crystal structure of LSD1-CoREST in complex with pLys4Met H3 peptide. (a) Ribbon diagram of the structure; LSD1 is in blue, CoREST in red, and the peptide in green. The FAD cofactor is shown as yellow ball-and-stick. The final model consists of residues 171-836 of LSD1, residues 308-440 of CoREST, and residues 1-16 of pLys4Met peptide. (b) Fitting of the refined pLys4Met peptide in the unbiased electron density calculated with weighted 2Fo-Fc coefficients. The map was calculated prior inclusion of the peptide atoms in the refinement calculations. The contour level is 1.2σ and the resolution is 3.1 Å. The sequence of the histone H3 peptide used for the X-ray analysis is: 1-ARTMQTARKSTGGKAPRKQLA-21.

Figure 2. Recognition of H3 peptide by LSD1. (a) Surface view of the peptide-binding pocket. The peptide is shown in green and the LSD1-surface in grey. Nitrogens are blue, oxygens red, sulphurs yellow, and carbons green. The positions of negatively charged residues lining the peptide-binding site are labeled. The Cα trace of CoREST residues 308-314 is shown in red highlighting their proximity to the LSD1 372-395 α-helix that is integral part of the peptide-binding site. The orientation is the same as in Figure 1a. (b) Three-dimensional view of the peptide-binding mode. Nitrogens are blue, oxygens red, and sulphurs yellow. Carbons of peptide and protein residues are in green and grey, respectively. The flavin cofactor is yellow. With respect to Figure 1a, the structure is rotated by approximately 180° about the vertical axis in the plane of the drawing. (c) Schematic representation of the peptide-protein interactions. (d) Modeling of di-methylLys4 peptide substrate bound in the active site. Orientation and atom colors are the same as in Figure 2b. The modeling was carried out assuming that the Cβ-Cγ-Cδ atoms of di-methylLys4 adopt the same conformation of the Cβ-Cγ-Sδ atoms of pMet4 in the crystal structure. In this way, the predicted position of the N-bound CH3 group of di-methylLys4 falls exactly in front of the N5 atom of the flavin. This type of substrate-binding geometry is similar to that found in other flavin-dependent oxidases and fully consistent with an oxidative attack of the flavin on the N-CH3 group of the substrate leading to the formation of an NH=CH2 imine that is then hydrolyzed to generate the demethylated Lys4 and formaldehyde.
Table 1. Kinetic parameters for LSD1 and LSD1-CoREST complex.

|                | $k_{cat}$ (min$^{-1}$)$^a$ | $K_m$ (μM)$^a$ | $K_i$ (μM)$^b$ |
|----------------|----------------------------|----------------|----------------|
| pLys4Met peptide |                            |                |                |
| LSD1           | 3.45 ± 0.04$^c$            | 3.42 ± 0.25$^c$| 0.04 ± 0.02    |
| LSD1-CoREST    | 7.35 ± 0.28               | 5.12 ± 1.04    | 0.05 ± 0.02    |

$^a$ Apparent steady-state kinetic parameters were determined as described by using a 21 amino acid H3 peptide monomethylated at Lys4 (7).

$^b$ The inhibition constant for the pLys4Met peptide was measured following previously described protocols (20).

$^c$ Data taken from (7).
Table 2. Data collection and refinement statistics

|                        | LSD1-CoREST complex with pLys4Met peptide |
|------------------------|-------------------------------------------|
| Space group            | I222                                       |
| Unit cell (Å)          | \(a = 120.1\) \(b = 180.5\) \(c = 233.4\) |
| Resolution (Å)         | 3.1                                        |
| \(R_{sym}^{a,b}\) (%)  | 10.9 (50.7)                                |
| Completeness\(^b\) (%) | 96.0 (98.1)                                |
| Unique reflections     | 44,088                                     |
| Redundancy             | 4.2 (4.1)                                  |
| \(I/\sigma\)\(^b\)    | 11.4 (2.3)                                 |
| \(N°\) of atoms        | protein/FAD/peptide\(^c\) 6346/53/114      |
| Average B value for    |                                             |
| ligand atoms (Å\(^2\)) | 64.2                                       |
| \(R_{crys}\)^\(d\) (%) | 22.3                                       |
| \(R_{free}\)^\(d\) (%) | 23.9                                       |
| Rms bond length (Å)    | 0.015                                      |
| Rms bond angles (°)    | 1.67                                       |

\(^a\) \(R_{sym} = \frac{\sum_i \sum_{h} |l(h)_{i} - \langle l(h) \rangle|}{\sum_i \sum_{h} l(h)_{i}}\), where \(l(h)_{i}\) is the scaled observed intensity of the \(i\)th symmetry-related observation for reflection \(h\) and \(\langle l(h) \rangle\) is the average intensity.

\(^b\) Values in parentheses are for reflections in the highest resolution shell.

\(^c\) The final model consists of residues 171-836 of LSD1, a FAD molecule, residues 308-440 of CoREST, and residues 1-16 of pLys4Met peptide.

\(^d\) \(R_{crys} = \frac{\sum F_{obs} - F_{calc}}{\sum |F_{obs}|}\), where \(F_{obs}\) and \(F_{calc}\) are the observed and calculated structure factor amplitudes, respectively. The set of reflections used for \(R_{free}\) calculations and excluded from refinement was extracted from the structure factor file relative to the PDB entry 2IW5 (12).
Figure 1a
Figure 1b
Figure 2b
Figure 2c
