PROTEIN STRUCTURE REPORTS

Crystal structure of the histone lysine specific demethylase LSD1 complexed with tetrahydrofolate

Zigmund Luka,[1]* Svetlana Pakhomova,[2] Lioudmila V. Loukachevitch,[1] M. Wade Calcutt,[1] Marcia E. Newcomer,[2] and Conrad Wagner[1]

[1]Department of Biochemistry, Vanderbilt University Medical Center, Nashville, Tennessee 37232
[2]Louisiana State University, Department of Biological Sciences, Baton Rouge, Louisiana 70803

Received 19 February 2014; Accepted 28 March 2014
DOI: 10.1002/pro.2469
Published online 8 April 2014 proteinscience.org

Abstract: An important epigenetic modification is the methylation/demethylation of histone lysine residues. The first histone demethylase to be discovered was a lysine-specific demethylase 1, LSD1, a flavin containing enzyme which carries out the demethylation of di- and monomethyllysine 4 in histone H3. The removed methyl groups are oxidized to formaldehyde. This reaction is similar to those performed by dimethylglycine dehydrogenase and sarcosine dehydrogenase, in which protein-bound tetrahydrofolate (THF) was proposed to serve as an acceptor of the generated formaldehyde. We showed earlier that LSD1 binds THF with high affinity which suggests its possible participation in the histone demethylation reaction. In the cell, LSD1 interacts with co-repressor for repressor element 1 silencing transcription factor (CoREST). In order to elucidate the role of folate in the demethylating reaction we solved the crystal structure of the LSD1–CoREST–THF complex. In the complex, the folate-binding site is located in the active center in close proximity to flavin adenine dinucleotide. This position of the folate suggests that the bound THF accepts the formaldehyde generated in the course of histone demethylation to form 5,10-methylene-THF. We also show the formation of 5,10-methylene-THF during the course of the enzymatic reaction in the presence of THF by mass spectrometry. Production of this form of folate could act to prevent accumulation of potentially toxic formaldehyde in the cell. These studies suggest that folate may play a role in the epigenetic control of gene expression in addition to its traditional role in the transfer of one-carbon units in metabolism.

Keywords: LSD1; tetrahydrofolate; crystal structure; epigenetics

Abbreviations: CoREST, co-repressor for repressor element 1 silencing transcription factor; diMeK4H3$_{1-21}$, dimethyl-Lys4 Histone H3 peptide aa1–21; DMGDH, dimethylglycine dehydrogenase; FAD, flavin adenine dinucleotide; LSD1, lysine specific histone demethylase 1; monoMeK4H3$_{1-21}$, monomethyl-Lys4 Histone H3 peptide aa1–21; SDH, sarcosine dehydrogenase; THF, tetrahydrofolate.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: U.S. DOE; Grant number: DE-AC02-06CH11357; Grant sponsor: Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor; Grant number: 085P1000817; Grant sponsor: NIH; Grant number: P30DK058404. Grant sponsor: National Institutes of Health; Grant numbers: DK15289 and DK080010 (to CW). Grant sponsor: Louisiana Governor's Biotechnology Initiative (to MEN).

*Correspondence to: Zigmund Luka, Vanderbilt University Medical Center, 604 Robinson Research Building, Nashville, TN 37232. E-mail: z.luka@vanderbilt.edu

Published by Wiley-Blackwell. © 2014 The Protein Society

Introduction

One of the most important recent discoveries in epigenetics was the identification of the enzymes that remove methyl groups, which serve as epigenetic markers, from lysine residues of histones.[1–5] Demethylation of mono- and dimethylated Lys4 on histone H3 is catalyzed by lysine specific demethylase 1 (LSD1). This enzyme is an amine oxidase containing flavin adenine dinucleotide (FAD) as the electron acceptor to first oxidize the lysine *N*-methyl amine to lysine *N*-methylimine. $FADH_2$ is reoxidized to FAD by molecular oxygen producing hydrogen peroxide. The *N*-methylimine is non-enzymatically hydrolyzed to a carbinolamine which spontaneously dissociates to a demethylated lysine and formaldehyde (Supporting Information Fig. S1). If the substrate is a dimethylated Lys4 the enzyme performs sequential removal of both methyl groups by the same mechanism. This mechanism is similar to demethylation of dimethylglycine by dimethylglycine dehydrogenase (DMGDH), sarcosine by sarcosine dehydrogenase (SDH), and other similar enzymes.[6,7] Both DMGDH and SDH bind tetrahydrofolate (THF). This was proposed to serve as a trap for toxic formaldehyde by formation of methylene-tetrahydrofolate (5,10-CH_2-THF) non-enzymatically. Based upon the similarity of reactions catalyzed by LSD1, DMGDH, and SDH, we proposed that LSD1 could also bind THF. In our previous publication we showed that, indeed, LSD1 binds THF with high affinity.[8]

In the cell, LSD1 demethylates histone H3 as part of a multimeric protein complex in which it directly interacts with the co-repressor of the repressor element 1 silencing transcription factor, CoREST. Interaction of these two proteins has been studied in detail and the crystal structure of the complex containing an analog of the histone H3 peptide was determined.[9,10] To elucidate the role of folate in histone H3 demethylation by LSD1 we have solved the crystal structure of the complex of LSD1–CoREST–THF and determined the folate binding site in LSD1.

Results

Folate-binding site

The overall structure of the LSD1–CoREST–THF complex as shown in Figure 1(A) is essentially the same as previously determined crystal structures of the LSD1–CoREST and LSD1–CoREST with a substrate-like peptide.[9,10] In the complex, LSD1 interacts with CoREST via its helix-tower without conformational changes in other protein domains. The THF-binding site in the crystal structure is located in the active center of LSD1 in close proximity to FAD in the amino oxidase domain.

Folate protein interactions and conformation of folate

Folates, in complexes with proteins, can exist in different (bent or extended) conformations depending on the specific folate–protein interactions.[11–13] In the LSD1–CoREST–THF complex, THF is observed in the bent conformation with the *p*-aminobenzoic (PABA) ring almost stacked to the pterin part of the molecule [Fig. 1(B)]. As in other folate-binding proteins, the THF is buried in the hydrophobic part of the protein. The pterin and PABA rings interact with the globular part of the protein. The glutamate part of folate is mostly oriented toward the surface of the protein. The hydrophobic interactions of THF with the protein are established through most of the atoms of the pterin and PABA rings with FAD, Val333, Phe538, Tyr761, and Ala809. THF is located 3.56 Å from the FAD. In addition to the non-specific hydrophobic interactions, THF establishes two hydrogen bonds with LSD1: between the N5 of the NH-group of the pterin ring and the carbonyl oxygen of Ala539 (3.09 Å) and the glutamate moiety (OX2) and ND1 of His564 (2.94 Å). There is also close separation (3.27 Å) between nitrogen N2 atom of the pterin ring and the OE1 oxygen of the glutamate moiety of the folate [Fig. 2(A), drawn by Ligplot[14]]. In summary, the position of THF in the active center of LSD1 is consistent with the proposed role of THF as an acceptor of the one-carbon unit derived from removed methyl group.

The position of THF in the active center of LSD1 suggests its direct participation in the demethylation reaction. It is not possible to determine the exact positions of THF and the natural peptide substrate in the active center without the reaction taking place. However, the crystal structure of LSD1–CoREST with an inactive mutated 21-amino acid N-terminal peptide of histone H3 in which Lys4 was substituted with methionine has been solved (ref. 10, PDB ID **2V1D**). Superposition of the folate binding site with this structure shows that folate occupies the site of the Met in the substrate analogue [Fig. 2(B)]. Because it is not possible for both THF and the substrate to bind in exactly the same position we believe that the structure determined by Forneris *et al.*[10] using the inactive substrate analogue is not the native conformation. The mutated peptide of histone H3 binds with the LSD1–CoREST complex with much higher affinity (a lower K_d) than the normal peptide[10] and should result in a slight change in conformation around the active site. This suggests that the natural Lys4-peptide is positioned in the active center in a different conformation than the mutated peptide and that THF and the natural peptide substrate are located in close proximity.

LSD1 activity in the presence of THF and 5,10-CH_2-THF formation

To check the possible role of THF as activator/inhibitor, we used mass spectrometry to monitor the disappearance of the diMeK4H3$_{1–21}$ substrate and the formation of two demethylation products,

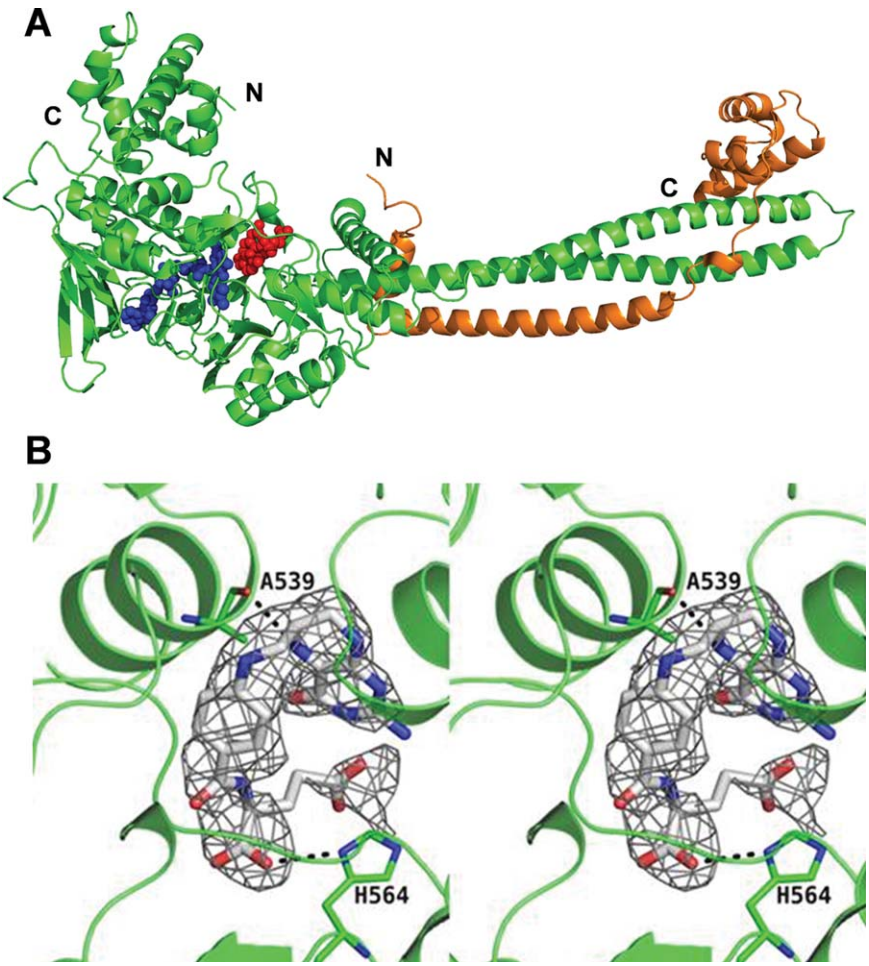

Figure 1. A: Overall structure of LSD1–CoREST–THF complex. LSD1 (green), CoREST (orange), FAD (blue), and THF (red) are drawn in Pymol. N- and C-termini of the proteins are marked. B: A stereo view of the folate binding site in the LSD1–CoREST complex. The $F_o - F_c$ electron density map is contoured at 3σ. The map was calculated before inclusion of the folate atoms into the model. Superimposed final model of the LSD1–CoREST complex is shown. Hydrogen bonds are indicated by dashes.

monoMeK4H3$_{1–21}$ and unmethylated K4H3$_{1–21}$ without and with THF in the reaction mixture. We found that at the concentration employed, THF had no effect on the time course of the concentrations of substrate, mono-methylated product, or completely demethylated product [Fig. 2(C)].

Use of mass-spectrometry allowed us to confirm the conversion of THF to 5,10-CH_2-THF during the demethylation reaction. Spectral data acquired at the 10 min reaction time point with and without THF [Fig. 2(D)] reveals an additional ion with $[M+H]^+$ m/z 458.2 in the presence of THF that corresponds to 5,10-CH_2-THF.

Discussion

In our work, we determined the position of the THF-binding site in LSD1 which provides the basis for understanding the role of THF in the histone H3 demethylation reaction. When studied *in vitro*, the reaction products formaldehyde and hydrogen peroxide are formed. These compounds are potentially toxic and should be neutralized by intracellular detoxification mechanisms.

The formaldehyde generated could be scavenged by THF which non-enzymatically reacts with formaldehyde and the product of this reaction is 5,10-CH_2-THF.[15,16] Therefore, if free formaldehyde is released in the course of histone demethylation it could be trapped by this non-enzymatic reaction with THF. In this case, THF should be bound to LSD1 in proximity to the active center. This is exactly what we found in the crystal structure of the LSD1–CoREST–THF complex.

The 5,10-CH_2-THF formed by reaction of formaldehyde with THF could be used for intracellular metabolism. As discussed previously,[8] one such pathway for thymidylate synthase exists in the nucleus in which 5,10-CH_2-THF is used for synthesis of deoxythymidine monophosphate (dTMP).[17] It would be reasonable to suggest that 5,10-CH_2-THF synthesized by oxidative demethylation of the histone H3 might be used as a substrate for thymidylate synthase as well.

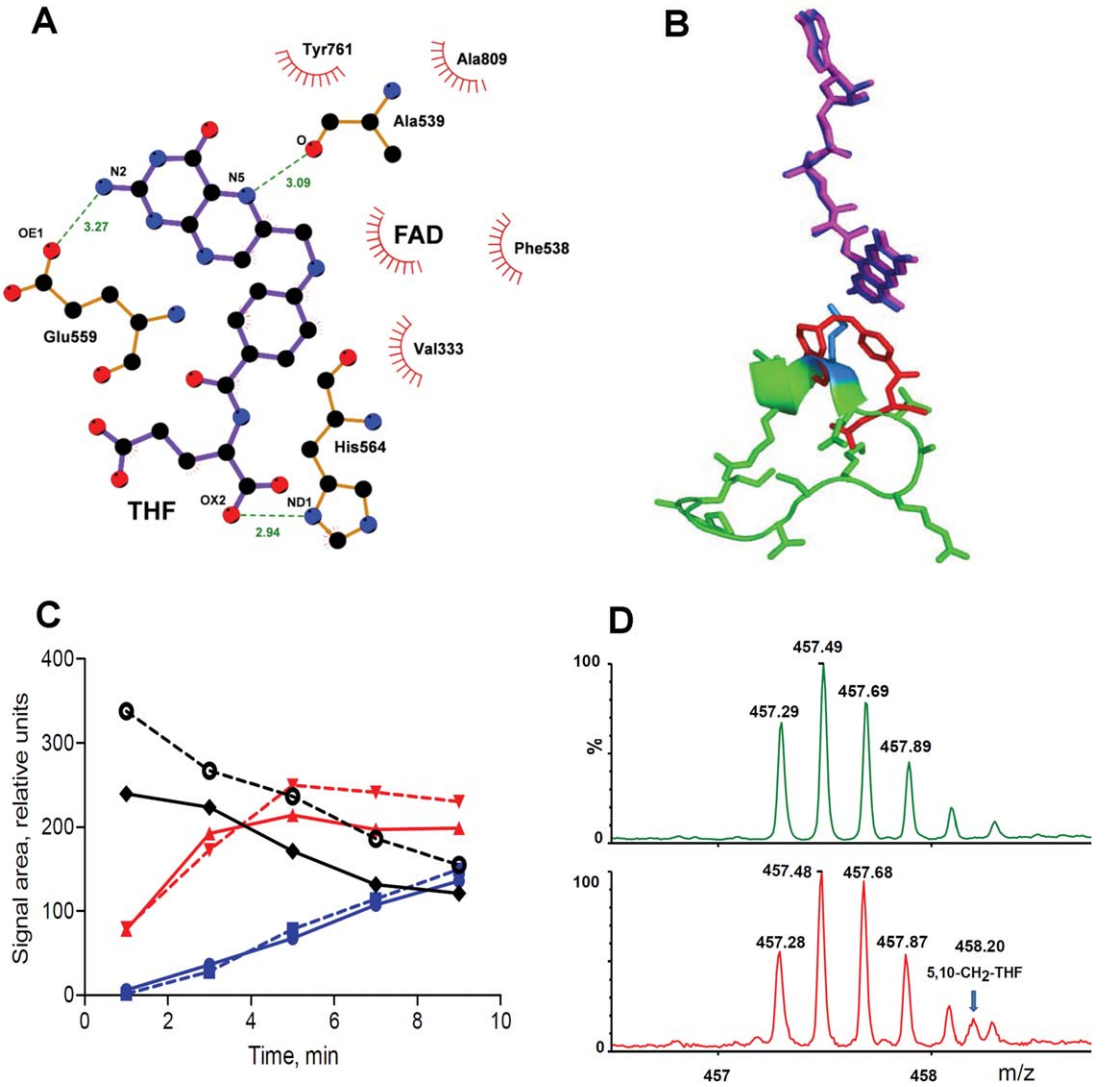

Figure 2. A: THF–protein interactions. Interactions of THF with LSD1 were analyzed by Ligplot and are presented in the software drawing. Atoms participating in hydrogen bonding are connected with dashed lines with the distance in Å. The protein residues participating in hydrophobic interaction are shown with Ligplot drawing for such interactions. B: Superposition of THF and substrate model in the LSD1–CoREST–THF complex. The crystal structure from this work and that from PDB ID 2V1D were superpositioned and folate (red), FAD molecules (blue and magenta), and model peptide (green with Met4 in blue) were drawn by Pymol. C: LSD1 activity and formation of 5,10-CH_2-THF in the course of demethylation reaction. Time course of LSD1 substrate diMeK4H3$_{1–21}$ (black lines) and two products of reaction, monoMeK4H3$_{1–21}$ (red lines) and unmethylated K4H3$_{1–21}$ (blue lines). The solid lines denote data for reaction without THF and dashed lines denote the data for reaction with THF. D: Formation of 5,10-CH_2-THF. On the upper panel the mass spectrum of the 5+ charge state of diMeK4H3$_{1–21}$ acquired for the 10 min time point of the reaction without (green line) and with (red line) THF in reaction mixture. Note an appearance of the ion with *m/z* 458.2, which corresponds to 5,10-CH_2-THF. The peaks with *m/z* 457 are 5+ charge state of diMeK4H3$_{1–21}$.

An important conclusion to be made from these results is the role that folate may play in the epigenetic control of gene expression. It suggests that this group of coenzymes has a significance beyond the traditional role as a carrier of one-carbon units.

Materials and Methods

All general chemicals and microbiological media were from Sigma. Dimethyl-Lys4 Histone H3 peptide aa 1–21 (diMeK4H3$_{1–21}$) was from BPS Bioscience. PreScission protease was from GE Healthcare. The natural (6*S*)-stereoisomer of tetrahydrofolate monoglutamate ((6*S*)-THF-Glu1) was a gift from EPROVA (Switzerland). The plasmids for the N-terminal truncated LSD1 (aa 171–852) and CoREST (aa 286–482 plus His-tag sequence) were a generous gift of Dr. Cole (Johns Hopkins University). The plasmid for full-size LSD1 expression was a generous gift of Dr. Shi (Harvard University).

The full size and truncated LSD1 were expressed in *E. coli* BL21(DE3) as we reported in the previous publication.[8] CoREST was expressed in *E. coli* BL21(DE3) in LB media with kanamycin (50 μg/mL) overnight at 19°C with 1 m*M* isopropy1-D-1-

thiogalactopyranoside (IPTG). Collected cells were sonicated in the homogenization buffer 0.1*M* Tris, pH 7.5, 2 μg/mL leupeptin, 2 μg/mL pepstatin, and 2 m*M* phenylmethylsulfonyl fluoride (PMSF) and crude extract was obtained by centrifugation. The crude extract was loaded onto column with Ni-NTA resin (QIAGEN), and non-bound proteins were washed-out with washing buffer containing 0.1*M* Tris, pH 7.5, 0.5*M* NaCl, 2 μg/mL leupeptin, 2 μg/mL pepstatin, 2 m*M* PMSF, and 20 m*M* imidazole. Bound proteins were eluted with elution buffer 50 m*M* Tris, pH 7.5, and 250 m*M* imidazole. Eluted proteins were concentrated and applied onto column with Whatman CM-52 cellulose equilibrated with 20 m*M* Tris, pH 7.6, 5 m*M* β-mercaptoethanol. After column washing with equilibration buffer, the CoREST was eluted by 20 m*M* Tris pH 7.5, 1*M* NaCl with purity greater than 95%.

Crystallization of LSD1–CoREST complex

The purpose of this work was to identify the folate-binding site in the LSD1. Our preliminary crystallization trial showed that the best way to do this was by using truncated LSD1 (which binds THF) complexed with CoREST. The LSD1–CoREST complex was prepared by mixing the LSD1 and CoREST stock solutions in a 1:1.5 molar ratio of LSD to CoREST. After incubating the mixture on ice for 1 h the solution conditions were changed to that used for crystallization 10–12 mg/mL proteins concentration in 25 m*M* 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES)-Na, pH 7.4, 100 m*M* NaCl, 5 m*M* DTT, 1 m*M* PMSF and the excess CoREST was removed by using an Amicon Ultra 50K concentrator.

The LSD1–CoREST complex was crystallized by the sitting drop method at room temperature in the conditions described earlier,[9] with small adjustment of the salts concentrations. The LSD1–CoREST complex was mixed with the reservoir solution containing 0.60*M* Li_2SO_4, 0.63*M* $(NH_4)_2SO_4$, 0.25*M* NaCl, 100 m*M* Na-citrate, pH 5.6, and 10 m*M* DTT. The crystals of the LSD1–CoREST complex appeared in one or two days and grew to the maximum dimensions in 4–6 days. The crystals belong to the orthorhombic *I222* space group with $a = 123.86$ Å, $b = 179.37$ Å, $c = 235.05$ Å.

The THF ligand was introduced into the crystals by the soaking method. The LSD1–CoREST complex crystals were incubated for 3 h in the cryoprotectant containing 0.76*M* Li_2SO_4, 0.74*M* $(NH_4)_2SO_4$, 0.35*M* NaCl, 100 m*M* Na-citrate, pH 5.6, 22 m*M* DTT, 23% (vol/vol) glycerol, and 10 m*M* THF. After soaking the crystals were flash-cooled in liquid nitrogen.

Data collection, structure solution, and refinement

X-ray diffraction data were collected at 100 K at LS-CAT beamline 21, Advanced Photon Source, Argonne National Laboratory, IL using a MAR 225 CCD detector. Data were processed and scaled using HKL2000 package.[18] Data collection and data processing statistics are summarized in Table I.

Table I. *Data Collection and Refinement Statistics*

Wavelength (Å)	0.97872
Resolution (Å)	3.05
Temperature (K)	100
Space group	*I222*
Cell dimensions	
a (Å)	123.86
b (Å)	179.37
c (Å)	235.05
Number of protein complexes per asymmetric unit	1
No. of unique reflections	47,051
R_{sym}[a,b] (%)	13.2 (67.3)
Completeness (%)	92.7 (76.1)
Redundancies	6.4 (4.1)
$I/\sigma(I)$	16.3 (1.9)
Refinement statistics	
Resolution range (Å)	40–3.05
No. of reflections used in refinement	44,261
σ cutoff used in refinement	none
R/R_{free}[c] (%)	19.59/21.73
Number of refined atoms	
Protein	6298
Heterogen atoms	92
Water	19
Average *B*-factors (Å^2)	
LSD1	43.6
CoREST	46.5
FAD	57.0
Folate	120.6
Water	52.0
R.m.s. deviations	
Bonds (Å)	0.011
Angles (°)	1.581
Ramachandran plot (%)	
Favored	92.3
Allowed	7.7
Generous	0
Disallowed	0

[a] Values in parentheses are for the highest-resolution shell.
[b] $R_{sym} = \sum|I_i - \langle I_i\rangle|/\sum I_i$, where I_i is the intensity of the ith observation and $\langle I_i\rangle$ is the mean intensity of the reflection.
[c] $R = \sum||F_o| - |F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factors amplitudes. R_{free} is calculated using 5.1% of reflections omitted from the refinement.

The molecular replacement procedure was applied to locate a solution using the program MOLREP.[19] The structure of the LSD1–CoREST complex (PDB accession code 2IW5) was used as a search model. The positioned MR model was refined using the maximum likelihood refinement in REFMAC with the TLS parameters generated by the TLSMD server.[19,20] The difference Fourier map revealed the presence of (6*S*)-THF in the active site of the complex [Fig. 1(B)]. The folate molecule was modeled according to the shape of electron density. Coot was

used for model building throughout the refinement.[21] The final model consists of residues 171–836 of LSD1, residues 308–440 of CoREST, one FAD molecule, one folate molecule, one chloride anion, one glycerol molecule, and 19 water molecules. Refinement statistics are listed in Table I.

Monitoring reaction time course and measuring formation of 5,10-CH_2-THF by mass-spectrometry

To monitor the reaction time course the substrate diMeK4H3$_{1–21}$ peptide was demethylated in a 100 μL reaction mixture containing 5 m*M* HEPES(Na), pH 7.5, 1.0 μ*M* of LSD1, 42 μ*M* of substrate, and 12.5 m*M* of DTT with or without 250 μ*M* of THF and levels of substrate and products analyzed by mass-spectrometry. Reaction mixture without substrate was pre-incubated with THF for 30 min at 25°C for protein-ligand binding when the effect of THF was studied.

Atomic Coordinates

The data were deposited to Protein Data Bank with PDB ID **4KUM**.

Acknowledgment

Authors thank Dr. Philip A. Cole and Dr. Yang Shi for the generous gifts of plasmids for LSD1 and CoREST expression. They also thank EPROVA for the generous gift of folates.

References

1. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119:941–953.
2. Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E (2005) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. FEBS Lett 579: 2203–2207.
3. Culhane JC, Cole PA (2007) LSD1 and the chemistry of histone demethylation. Curr Opin Chem Biol 11:561–568.
4. Anand R, Marmorstein R (2007) Structure and mechanism of lysine-specific demethylase enzymes. J Biol Chem 282:35425–35429.
5. Lan F, Nottke AC, Shi Y (2008) Mechanisms involved in the regulation of histone lysine demethylases. Curr Opin Cell Biol 20:316–325.
6. Wittwer AJ, Wagner C (1981) Identification of the folate-binding proteins of rat liver mitochondria as dimethylglycine dehydrogenase and sarcosine dehydrogenase. Purification and folate-binding characteristics. J Biol Chem 256:4102–4108.
7. Wittwer AJ, Wagner C (1981) Identification of the folate-binding proteins of rat liver mitochondria as dimethylglycine dehydrogenase and sarcosine dehydrogenase. Flavoprotein nature and enzymatic properties of the purified proteins. J Biol Chem 256:4109–4115.
8. Luka Z, Moss F, Loukachevitch LV, Bornhop DJ, Wagner C (2011) Histone demethylase LSD1 is a folate-binding protein. Biochemistry 50:4750–4756.
9. Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H (2006) Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell 23:377–387.
10. Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A (2007) Structural basis of LSD1-CoREST selectivity in histone H3 recognition. J Biol Chem 282:20070–20074.
11. Poe M, Benkovic SJ (1980) 5-Formyl- and 10-formyl-5,6,7,8-tetrahydrofolate. Conformation of the tetrahydropyrazine ring and formyl group in solution. Biochemistry 19:4576–4582.
12. Matthews DA, Alden RA, Bolin JT, Freer ST, Hamlin R, Xuong N, Kraut J, Poe M, Williams M, Hoogsteen K (1977) Dihydrofolate reductase: X-ray structure of the binary complex with methotrexate. Science 197:452–455.
13. Luka Z, Pakhomova S, Loukachevitch LV, Egli M, Newcomer ME, Wagner C (2006) 5-Methyltetrahydrofolate is bound in intersubunit areas of rat liver folate-binding protein glycine *N*-methyltransferase. J Biol Chem 282:4069–4075.
14. Wallace AC, Laskowski RA, Thornton JM (1996) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8:127–134.
15. Blakley RL (1958) Interaction of formaldehyde and tetrahydrofolic acid and its relation to the enzymic synthesis of serine. Nature 182:1719–1722.
16. Kallen RG, Jencks WP (1966) The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. J Biol Chem 241:5851–5863.
17. Stover PJ, Field MS (2011) Trafficking of intracellular folates. Adv Nutr 2:325–331.
18. Otwinowski Z, Minor W, Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW Jr, Sweet RM, Eds. (1997) Methods in enzymology, Vol. 276: Macromolecular crystallography, Part A. New York: Academic Press, pp 307–326.
19. Bailey S (1994) The Ccp4 suite—programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
20. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. J Appl Crystallogr 39:109–111.
21. Emsley P, Lohkamp B, Scott W, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501.