An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1-α

Karishma L. Manzur[1], Ming-Ming Zhou*

Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, 1425 Madison Avenue, New York, NY 10029, USA

Received 12 April 2005; revised 4 May 2005; accepted 11 May 2005

Available online 4 June 2005

Edited by Frances Shannon

Abstract **The evolutionarily conserved SET domain proteins in eukaryotes have been shown to function as site-specific histone lysine methyltransferases, and play an important role in regulating chromatin-mediated gene transcriptional activation and silencing. Structure-based sequence analysis has revealed that SET domains are also encoded by viruses and bacteria, as well as Archaea. However, their cellular functions remain elusive. In this study, we have characterized a SET domain protein from *Methanosarcina mazei* strain Gö1 that we refer to as Gö1-SET. We show that Gö1-SET exists as a homodimer in solution, and functions as a lysine methyltransferase with high substrate specificity that is dependent on the amino acid sequence flanking the lysine methylation site. Particularly, Gö1-SET exhibits selective methyltransferase activity towards one of the major archaeal DNA interacting protein MC1-α at lysine 37. Our findings suggest that SET domain proteins such as Gö1-SET may restructure archaeal chromatin that is composed of MC1–DNA complexes, and that modulation of chromatin structure by lysine methylation may have arisen before the divergence of the archaeal and eukaryotic lineages.**

Keywords: SET domain; Methyltransferase activity; *Methanosarcina*

1. Introduction

On the basis of their molecular properties, cellular life is classified into three domains – Archaea, bacteria and eukarya. Archaea lack nuclear membranes and are therefore prokaryotes, however, they are genetically and biochemically as divergent from bacteria as are eukarya. One of the main distinguishing features between the members of the three kingdoms of life lies in the mechanisms underlying gene expression. Archaea and bacteria share the same basic mechanism of translation, which differs from that used by the eukarya mainly at the initiation step involving ribosome binding to mRNA. On the other hand, the basal transcriptional machineries of Archaea and eukarya are related especially with respect to the structure of the core of the RNA polymerase as well as basic transcription factors. However, transcriptional regulation in Archaea most often follows mechanisms involving simple bacteria-like regulators, with certain occurrences of eukarya-like regulators [1]. By contrast, in eukarya the modulation of chromatin structure is far more complex involving covalent modification of DNA-associated histone proteins, which plays a pivotal role in transcriptional regulation [2].

Eukaryotic chromosomal DNA is structured by interaction with histones forming higher order DNA–protein complexes termed chromatin. The fundamental unit of eukaryotic chromatin is the nucleosome, made up of 146 bp of DNA wrapped around a histone octamer core [3]. Histones are small basic proteins consisting of a globular domain and flexible and charged N- and C-terminal tails that protrude from the nucleosome and serve as targets for post-translational modifications. Reversible covalent modifications on histones include acetylation, methylation and phosphorylation, which affect the electrostatic interaction of the histone–DNA complex, thereby leading to changes in the local chromatin structure, with subsequent effects on various cellular processes, including transcriptional regulation (reviewed in [4–7]).

Study of DNA packaging in Archaea reveals that there are no conserved mechanisms for archaeal genome packaging and that the processes are unique, with limited similarities to the eukaryotic and bacterial mechanisms [8]. Archaea contain a variety of sequence-independent DNA-binding proteins some of which are known to undergo post-translational modifications, similar to the histone modifications in eukaryotic chromatin. Among these archaeal DNA-binding proteins are Alba, which is conserved among most of the sequenced archaeal genomes [9,10], the so-called histone-like proteins in the subdomain Euryarchaea [8], and MC1 proteins in the *Methanosarcina* species [8].

To date, Alba has been the most well characterized among the known archaeal DNA-associating proteins. Alba proteins interact with DNA as dimers and such binding introduces negative supercoiling of the bound DNA [9,11]. Deacetylation of Alba by archaeal Sir2-like NAD-dependent protein deacetylases has been shown to increase the binding affinity of Alba for DNA [10]. Furthermore, deacetylation of Alba has been found to inhibit transcription in vitro [10]. Such studies indicate that similar to the eukaryotic histone proteins, archaeal DNA-associating proteins are also subject to post-translational modifications with subsequent effects on chromatin structure and function.

To further investigate the presence of archaeal homologs of eukaryotic histone-modifying proteins, we conducted a sequence homology-based search of the protein database using

*Corresponding author. Fax: +1 212 860 3369.
E-mail address: ming-ming.zhou@mssm.edu (M.-M. Zhou).

[1] Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

0014-5793/$30.00
doi:10.1016/j.febslet.2005.05.026

BLAST [12]. Based on our earlier studies on histone lysine methyltransferases, we primarily focused on identifying archaeal homologs of such eukaryotic methyltransferases. Histone lysine methylation in eukaryotes is catalyzed almost exclusively by the evolutionarily conserved SET domain [13,14], which was originally identified in *Drosophila* proteins: Suppressor of variegation (Su(var)3-9) [15], Enhancer of zeste (E(z)) [16,17] (hence the name). Similar to Aravind and Iyer [18], our search identified a SET domain encoding gene from the acetate-utilizing archaeal methanogen, *Methanosarcina mazei* strain Gö1. In addition, our search identified a SET encoding gene from another archaeal methanogen, *M. acetivorans* strain C2A.

The archaeon *Methanosarcina* and related species are obligate anaerobes of immense ecological importance. Acetate-utilizing archaeal methanogens are indispensable members of the anaerobic food chain since they are the only organisms capable of fermenting acetate, methylamines, and methanol to methane, carbon dioxide and ammonia [19]. It was intriguing to identify SET proteins in Archaea and therefore, to establish the functional significance of such proteins, we selected the SET protein from *M. mazei* strain Gö1 (which we refer to as Gö1-SET) for characterization as the first archaeal SET protein.

2. Materials and methods

2.1. Construction of expression vectors

Most archaeal genes do not contain introns. Therefore, it was possible to clone the following genes of interest from the genomic DNA of *M. mazei* strain Gö1 (gift from Gerhard Gottschalk, Georg-August-University, Göttingen, Germany). Gö1-SET (Accession No. NC_003901) was amplified from genomic DNA using the primers 5′-ATGCAGGAGCCTGATCAAATGTCATTG and 5′-GTTCTGGCTTGCAACATCAAACCAG. For expression of recombinant Gö1-SET protein without an epitope tag, the resulting DNA fragment was inserted into *Nde*I/*Bam*HI sites of pET-22b(+) vector (Novagen) with a stop codon inserted at the 3′-end.

Histone-like protein (Accession No. NP_633849) was amplified from genomic DNA using the primers 5′-ATGGCAGCAAAAGTTATACCGTTCGC and 5′-CAGCATTTCTTTAGCTAGTTTTATG. MC1-α (Accession No. NP_633075) was amplified from genomic DNA using the primers 5′-ATGGCTGACACACGAAATTTTGTTTTGCGAGA and 5′-CTCGATTGTTTCTATTTTTTCTTTCTTAACA. MC1-β (Accession No. NP_633245) was amplified from genomic DNA using the primers 5′-ATGTCCAACACAAGAAATTTTGTTTTACGAGAC and 5′-CTCGAGTTTTTCGATCTTTTCTTTCTTAACG. For expression of recombinant protein fused to an N-terminal hexahistidine tag, the resulting *Nde*I/*Bam*HI digested DNA fragments were inserted into the respective sites of pET-15b vector (Novagen).

Mutations were introduced into the DNA constructs using the QuikChange site-directed mutagenesis kit (Stratagene). All constructs were verified by DNA sequencing.

2.2. Protein sample preparation

DNA constructs were expressed in *E. coli* BL21(DE3) cells (Novagen). A uniformly ^{15}N-labeled protein sample was prepared by growing bacteria in a minimal medium containing $^{15}NH_4$ Cl as the sole nitrogen source. For NMR-based experiments, untagged Gö1-SET was expressed, refolded and purified as previously described [20]. Gö1-SET was further purified via gel filtration chromatography using Superdex 200 16/60 (Amersham Biosciences) at 1 ml min^{-1} in 50 mM phosphate, pH 6.5, buffer containing 700 mM NaCl, 300 mM urea, 0.1 mM EDTA, and 5 mM β-ME. Hexahistidine-tagged proteins were purified with Ni-NTA agarose beads according to the manufacturer's specifications (Qiagen).

2.3. Histone methyltransferase assay and preparation of its substrates

The in vitro histone methyltransferase assays were carried out as previously described [20]. Briefly, the enzymatic reactions were set-up using 0.5 μg of enzyme, 5 μg of substrates and 75 nCi of ^{14}C-labeled *S*-adenosyl-L-methionine (Amersham Biosciences) in 50 μl total reaction volume and incubated for 1 h at 37 °C. The buffer condition was 20 mM Tris–HCl, pH 8.0, containing 20 mM KCl, 10 mM $MgCl_2$, and 10 mM β-ME. Commercially available free core bovine histones H2A, H2B, H3 and H4 were also used as potential substrates in the methyltransferase assays (Roche). Resulting protein products from the methylation reactions were separated by SDS–PAGE and visualized by Coomassie staining and fluorography (Molecular Dynamics). Incorporation of ^{14}C-methyl groups onto synthetic histone peptide substrates (Bio-synthesis) were assayed in three independent experiments and detected by liquid scintillation using Ready Protein cocktail (Beckman Coulter).

2.4. NMR spectroscopy

NMR spectra were acquired at 37 °C on a 500 MHz Bruker DRX NMR spectrometer. ^{1}H and ^{15}N spectra were recorded on uniformly ^{1}H/^{15}N-labeled protein samples. NMR spectra were processed using NMRView [21].

3. Results

3.1. SET domain proteins are present in archaeal methanogens

Sequence homology-based search of the non-redundant database of protein sequences (NCBI) was conducted using the BLAST program similar to the evolutionary analysis of SET proteins conducted by Aravind and Iyer [12,18]. These authors identified SET domain encoding genes in various viruses, bacteria, and eukaryota, as well as the acetate-utilizing archaeal methanogen, *M. mazei* strain Gö1 [18]. In addition, our search identified a SET encoding gene from another archaeal methanogen, *M. acetivorans* strain C2A. We have, respectively, termed the archaeal gene products as Gö1-SET and C2A-SET. The identified open-reading frames encode for a 134-residue Gö1-SET and a 128-residue C2A-SET.

Primary sequence analysis of the archaeal SET proteins shows high sequence similarity with a viral SET protein called vSET, which we have previously characterized [20]. The 119-residue vSET consists of a single SET domain without any flanking sequences, thus representing a prototype of the SET domain protein family. Interestingly, all of the bacterial and archaeal SET domains identified in our search also appear to be full-length proteins constituting a single SET domain, similar to vSET.

Sequence alignment of vSET, Gö1-SET and C2A-SET, along with several bacterial SET proteins, was composed based on the three-dimensional structure of vSET [20], is presented in Fig. 1. The sequence alignment shows that residues that are important for histone methyltransferase activity of SET proteins, corresponding to Gly17, Asn69, His70, Asn75, Glu100, and Tyr105 in vSET, are also conserved in the two archaeal SET proteins.

3.2. Gö1-SET appears to exist as a dimer in solution

Our previous studies of vSET showed that the protein exists as a homodimer in solution, and we were able to characterize the residues present at the dimer interface from our three-dimensional structural analysis of vSET by NMR [20]. The primary sequence alignment in Fig. 1 shows that some of the residues known to be present at the dimer interface of vSET