Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3

Shumin Wu\textsuperscript{a}, Raymond C. Trievel\textsuperscript{b}, Judd C. Rice\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} University of Southern California Keck School of Medicine, Department of Biochemistry and Molecular Biology, 1501 San Pablo Street, ZNI 225, MC 2821, Los Angeles, CA 90033, United States

\textsuperscript{b} University of Michigan, Department of Biological Chemistry, Ann Arbor, MI 48109-0606, United States

Received 11 April 2007; revised 7 June 2007; accepted 14 June 2007

Available online 21 June 2007

Edited by Frances Shannon

Abstract Human SFMBT (hSFMBT) is postulated to be a Polycomb (PcG) protein. Similar to other PcG proteins, we found that hSFMBT displays robust transcriptional repressor activity. In addition, hSFMBT localized to the nucleus where it strongly associates with chromatin by directly and selectively binding the N-terminal tail of histone H3. Importantly, we discovered that the four tandem MBT repeats of hSFMBT were sufficient for nuclear matrix-association, N-terminal tail H3 binding, and required for transcriptional repression. These findings indicate that the tandem MBT repeats form a functional structure required for biological activity of hSFMBT and predict similar properties for other MBT domain-containing proteins.

Keywords: SFMBT; Polycomb; Transcription; Chromatin; Histone H3

1. Introduction

The ability to regulate the transcription of specific sets of genes is critical in determining cell fate. Importantly, these transcription patterns are propagated to progeny by epigenetic mechanisms that ensure the maintenance of cellular identity. Gene expression of developmental-associated genes is largely regulated by multimeric protein complexes that function to activate or repress transcription, both of which must operate in concert for proper differentiation. The best studied of these are regulatory complexes that contain the Polycomb group (PcG) or Trithorax (TRX) proteins \cite{1}. The PcG proteins, which were first discovered in \textit{Drosophila}, function to specify positional identity by creating a repressive chromatin structure at homeotic (Hox) genes resulting in their transcriptional silencing \cite{2,3}. Currently there are three known PcG complexes: PRC1, PRC2 and the newly characterized PhoRC \cite{4}. The main component of the PhoRC complex is Pleiohormetic (Pho), a sequence-specific DNA-binding protein that targets Polycomb response elements (PREs) in the genome \cite{5}. \textit{Drosophila} Pho was recently shown to heterodimerize with a novel PcG protein, known as SFMBT (Scm-related gene containing four mbt domains); which is required for \textit{Hox} gene silencing \cite{6}.

The mammalian version of SFMBT was first cloned seven years ago, however, little is currently known about its biological function \cite{7}. The translated protein contains four tandem malignant brain tumor (MBT) domains and a conserved protein-interacting sterile alpha motif (SAM) domain, which was first identified in the PcG gene Scm and is also found in both \textit{ph} and \textit{l(3)mbt} \cite{8}. The MBT domain is evolutionarily restricted to metazoan lineages, is invariably found in tandem arrays of two to four repeats and proteins harboring MBT domains, such as \textit{Drosophila} Sfmbt (dSfmbt), human sex comb on midleg-like 2 (SCML2) and lethal (3) malignant brain tumor (L(3)MBT), have been linked to \textit{PcG} silencing, although their function in these pathways remains elusive \cite{9,10}. Outside of these observations, little else is known about the mammalian homologs of SFMBT.

To gain further insights into the biological significance of human SFMBT (hSFMBT), we investigated the structure and function of this protein. Consistent with its role as a putative PcG protein, hSFMBT specifically partitions to the nucleus and is a potent repressor of transcription. We discovered that hSFMBT strongly interacts with the nuclear matrix and that it also selectively binds histones H3 and H4, both in vitro and in vivo. Binding occurs at the N-terminal tail suggesting that hSFMBT functions to sequester transcriptionally inert chromatin at the nuclear periphery. Interestingly, we discovered that all four of the MBT repeats of hSFMBT were sufficient and necessary for nuclear matrix attachment, transcriptional repression and histone binding. In addition, all four MBT domains were required for repressor activity indicating that the higher-order structure formed by the four MBT repeats is essential for biological function. This is consistent with the structural characterization of the MBT repeats in human SCML2 and L(3)MBT, where the MBT domains fold cooperatively through interdigitation to form unique higher-order structures \cite{9,11}. Lastly, we found that hSFMBT is preferentially expressed in certain cell types suggesting that it may be an important regulator of transcriptional programs during developmental and differentiation processes.

2. Materials and methods

2.1. Cell culture

HeLa, HEK-293 and K562 cells (ATCC) were cultured as previously described \cite{12}.

Abbreviations: SFMBT; Scm-related gene containing four mbt domains; dSfmbt, \textit{Drosophila} SFMBT; hSFMBT, human SFMBT; MBT, malignant brain tumor; SAM, sterile alpha motif; L(3)MBT, lethal (3) malignant brain tumor; SCML2, sex comb on midleg-like 2; PcG, Polycomb
2. Plasmids

The human full length SFMBT cDNA clone (MGC:3342 IMAGE:3029598), 4x MBT (aa 20–453), ΔMBT3.4 (aa 20–235), ΔMBT1.4 (aa 125–351), ΔMBT1.2 (aa 235–453), ΔMBT4 (aa 20–351) and ΔMBT1 (aa 125–453) were produced by PCR amplification and inserted in-frame in the pCMX-Gal4-DBD plasmid [13], the pGEX-4T-1 plasmid (GE Healthcare) or pEGFP-C1 plasmid (Clontech).

2.3. Microscopy

Microscopy of HeLa cells was performed as previously described [12].

2.4. Antibodies

A synthetic peptide corresponding to amino acids 767–780 of human SFMBT (NP_057413) was used to immunize rabbits under the Standard Protocol (Zymed Laboratories). A 1:100 dilution of sera was used for Western analysis and peptide competition experiments, as previously described [12]. The GFP antibody (ABCAM) was used 1:100,000, HP1β (Chemicon) was used 1:5000, RNAP II (Covance) was used 1:100,000, Gal4-DBD (Santa Cruz) was used 1:5000, H3 (ABCAM) and H4 (Upstate) were used 1:5000, Ub69 (Santa Cruz) was used 1:100, monoclonal FLAG (SIGMA) was used 1:10000 and the GST antibody (Upstate) was used 1:500 for Western analysis. For immunoprecipitations, 20 μl of the Gal4-DBD antibody or 20 μl M2 FLAG beads (SIGMA) were used.

2.5. Transfections and reporter assays

HEK-293 cells were plated in 6-well plates (4 x 10^5 cells/ml) and co-transfected 24 h later using Lipofectamine with Plus reagent (Invitrogen) and pCMX-Gal4-DBD fusion plasmid, 1 μg of reporter vector (pGK1-luc [14] or pSV40-luc [15]), 5 μg of pRL (Promega) and the pUC19 plasmid to bring the total DNA amount to 2 μg, according to the manufacturer’s instructions. Cell lysates were collected 48 h post-transfection and luciferase activity was measured using Luciferase Assay Substrate (Promega) and a TopCount NMR microplate reader (Packard Bioscience). Quantitative measurements were obtained by normalizing to renilla luciferase activity. Standard error bars were generated by performing all experiments in triplicate.

2.6. Chromatin fractionation

Isolation of the S1, S2 and P fractions were performed as previously described with minor modifications [16]. Nuclei from 2 x 10^6 cells were isolated and resuspended in 300 μl nuclear buffer (20 mM Tris–HCl, pH 7.5, 70 mM NaCl, 20 mM KCl, 5 mM MgCl2, 3 mM CaCl2, and protease inhibitors), split equally into 4 aliquots and incubated with 6 μg of monoclonal beads overnight at 4°C. The supernatant (S1) was collected and placed on ice while the nuclear pellet was lysed in 2 mM EDTA for 15 min at 4°C. The supernatant (S2) was collected following centrifugation and the pellet (P) was resuspended in lysis buffer (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 5 mM EDTA, 0.5 mM SDS). Western analysis of the fractions was performed with the indicated antibodies.

2.7. GST pull-downs

pGEX-4T-1, pGEX-4T-1-SFMBT and pGEX-4T-1-4x MBT were transformed into BL21 E. coli (Novagen). Expression of recombinant GST fusion proteins was induced using 0.5 mM IPTG (Calbiochem) and 100 μM DTT, 1% Triton X-100, protease inhibitors) and rotated for 30 min at 4°C. Lysates were clarified by centrifugation at 18000 x g for 1 min. The supernatant was dialyzed in PBS for 1 h at 4°C and incubated with either 20 μl Gal4-DBD antibody or 20 μl M2 FLAG-conjugated beads overnight at 4°C. Beads were washed with PBS, the bound material was eluted by boiling in SDS loading buffer and fractionated by SDS–PAGE prior to Western analysis.

2.8. Immunoprecipitations

HEK-293 cells were collected 24 h post-transfection, washed in PBS, resuspended in 300 μl lysis buffer (50 mM Tris pH 7.0, 150 mM NaCl, 0.5 mM DTT, 1% Triton X-100, protease inhibitors) and rotated for 30 min at 4°C. Lysates were clarified by centrifugation at 18000 x g for 1 min. The supernatant was dialyzed in PBS for 1 h at 4°C and incubated with either 20 μl Gal4-DBD antibody or 20 μl M2 FLAG-conjugated beads overnight at 4°C. Beads were washed with PBS, the bound material was eluted by boiling in SDS loading buffer and fractionated by SDS–PAGE prior to Western analysis.

3. Results

3.1. SFMBT is a conserved Polycomb-like protein

In order to gain insights into the possible biological functions of hSFMBT, we compared its amino acid composition and structure to other metazoan MBT-containing proteins. The human, mouse and rat SFMBT proteins are structurally similar, where each contains four N-terminal tandem MBT repeats, a SAM domain near the C-terminus. While Drosophila SFMBT (dSfmbt) retains the C-terminal SAM domain, its four tandem MBT repeats are located towards the C-terminal and dSfmbt contains a zinc finger motif which is lacking in mammals (not shown). Despite these structural differences, a sequence alignment of the MBT domains of dSfmbt, hSFMBT, and hL3(3)MBT reveals a substantial degree of homology within their respective repeats (Fig. 1). A pairwise alignment of the individual repeats from the fly and human SFMBT proteins revealed that there is 48% sequence homology and 34% sequence identity between the two. This high degree of conservation strongly suggests that hSFMBT functions as a Polycomb-group protein, similar to dSfmbt [6].

3.2. hSFMBT is a cell type-specific nuclear protein

Since hSFMBT is a putative PcG protein, we hypothesized that it would function as a nuclear protein. To determine this, a GFP plasmid fused in-frame to full length wild type hSFMBT was created and transiently transfected into HeLa cells. Fluorescence microscopy of the cells revealed that GFP–SFMBT was enriched within nuclei as demonstrated by its co-localization with nuclear DAPI staining (Fig. 2A). In contrast, cells transfected with a GFP construct lacking an insert was evenly dispersed throughout the HeLa cells. The transfected HeLa cells were also fractionated into nuclear and cytoplasmic components and Western analysis of the fractions using a GFP antibody confirmed that the GFP–SFMBT fusion protein was specifically enriched within the nuclear compartment (Fig. 2B). To verify that endogenous hSFMBT partitioned to the nucleus, a novel polyclonal SFMBT-specific antibody was created (Supplemental Fig. 1) and used in Western analysis of whole cell lysates from several commonly used human cell lines. Interestingly, hSFMBT was only detected in specific cell types, mainly those of hematological origin (Fig. 2C). While the highest levels of hSFMBT were detected in the erythroblastic K562 and myeloblastic HL-60 cells, hSFMBT was also detected in the B-cell lymphoblastic Daudi cells. In contrast, hSFMBT was barely detected in epithelial cell lines derived from uterine (HeLa), breast (MCF7) and kidney (HEK-293) tissues. Since the K562 cells expressed the highest levels of hSFMBT, they were fractionated into nuclear and cytoplasmic components for Western analysis with the SFMBT antibody.
Similar to the GFP–SFMBT fusion protein, we found that endogenous hSFMBT was also specifically enriched within the nuclear compartment (Fig. 2D). Collectively, these findings demonstrate that hSFMBT is a cell type-specific nuclear protein.

3.3. hSFMBT strongly associates with the nuclear matrix

To further dissect the sub-nuclear localization of hSFMBT, nuclei were isolated from K562 cells, as depicted in Fig. 3A, and were partially digested with micrococcal nuclease (MNase) for 1, 4 or 16 min before isolating the various chromatin components by centrifugation [16,18]. DNA analysis demonstrates that the MNase-sensitive soluble S1 fraction is mainly composed of mono- and dinucleosomal sized DNA fragments, typically associated with euchromatin (Fig. 3B). In contrast, the insoluble S2 fraction is composed of MNase-resistant oligonucleosomes, typically associated with heterochromatin, as observed by the higher molecular weight laddering. The P fraction represents the nuclear material that remains bound to the nuclear matrix. With increased MNase digestion time, more of the S2 and P fractions become soluble and shift into the S1 and S2 fractions, respectively. Western analysis of these fractions using the hSFMBT antibody revealed that endogenous hSFMBT selectively and strongly associates with the nuclear matrix.
matrix as it failed to shift from the P fraction even at extended MNase digestion times (Fig. 3C). Similarly, RNAP II was found to preferentially associate with the nuclear matrix, as previously reported [19]. In contrast, the beta isoform of heterochromatin protein 1 (HP1β) was enriched in the S2 fraction but was liberated to the S1 fraction upon extended digestion with MNase [20]. These findings demonstrate that hSFMBT strongly interacts with the nuclear matrix.

Due to the evolutionary conservation of the four tandem MBT domains (Fig. 1), we hypothesized that the MBT repeats were sufficient for nuclear matrix attachment. To test this, a Gal4–DBD fusion construct with an hSFMBT truncation mutant containing only the four MBT repeats (4x MBT) was transfected into HEK-293 cells for analysis (Fig. 3A). Similar to the findings for endogenous hSFMBT, the 4x MBT fusion protein was preferentially bound to the nuclear matrix while the Gal4–DBD-null plasmid was ubiquitously distributed amongst the fractions (Fig. 3D). Therefore, these findings indicate that the four MBT repeats of hSFMBT are sufficient for nuclear matrix attachment.

3.4. The four MBT repeats of hSFMBT are sufficient for potent transcriptional repression

Since SFMBT associates with the nuclear matrix, similar to RNAP II, we hypothesized that SFMBT could function as a co-activator of transcription. To test this hypothesis, the pGK1-luc reporter construct containing five tandem repeats of the Gal4 upstream activating sequence (5x UAS) followed by a TATA box and luciferase reporter gene was employed as previously described (Fig. 4A) [21]. HEK-293 cells were co-transfected with pGK1-luc and Gal4–DBD fusion constructs containing either the CARM1 co-activator as the positive control [22], full length hSFMBT or 4x MBT; a Gal4–DBD-null vector served as the negative control. In addition, cells were co-transfected with the pRL reporter vector to normalize for transfection efficiency [23]. As predicted, the CARM1 co-activator increased luciferase gene expression by sevenfold compared to control (Fig. 4A). In contrast, both Gal4–DBD-SFMBT and Gal4–DBD-4x MBT failed to activate transcription but, instead, greatly reduced even basal levels of luciferase expression when compared to the Gal4–DBD-null negative control. These findings demonstrate that hSFMBT does not act as a co-activator of transcription and, rather, suggests that hSFMBT may function as a transcriptional repressor protein.

To test this hypothesis, the pSV40-luc reporter construct containing five tandem repeats of the Gal4 upstream activating sequence (5x UAS) followed by an SV40 promoter that constitutively activates robust transcription of a luciferase reporter gene was employed as previously described (Fig. 4B) [24]. HEK-293 cells were co-transfected with the pSV40-luc reporter and the Gal4–DBD fusion constructs generated above. The Gal4–DBD-
SMRT repressor protein served as the positive control for repression [25] and the pRL reporter vector was also used to normalize for transfection efficiency. Consistent with previous reports, the SMRT repressor protein produced a fivefold decrease in transcription of the luciferase gene compared to the negative control (Fig. 4 B). Interestingly, both Gal4–DBD-SFMBT and Gal4–DBD-4\cdot MBT produced a more than sevenfold and ninefold decrease in luciferase gene expression, respectively. The failure of a FLAG-tagged full length SFMBT construct to reduce luciferase activity in these assays confirmed that the observed repressive effects of Gal4–DBD-SFMBT were not indirectly due to ectopically expressed hSFMBT. These data indicate that hSFMBT is a potent repressor of transcription and that the four MBT repeats of SFMBT are sufficient to induce repression.

3.5. All four MBT repeats of hSFMBT are required for repressor function

To further define which of the MBT repeats were required for the observed repressive effects of hSFMBT, truncation mutants of the Gal4–DBD-4\cdot MBT fusion construct were created, as depicted in Fig. 4C, and used in the repression assays. Initial studies of constructs lacking the two MBT repeats closest to the N-terminal (ΔMBT1,2) or C-terminal (ΔMBT3,4) both resulted in a complete loss of repression. Therefore, we speculated that the two central MBT repeats were required for repression, however, these two repeats (ΔMBT1,4) alone were not sufficient to restore repression suggesting that an additional MBT repeat flanking either the N-terminal (ΔMBT4) or C-terminal (ΔMBT1) was also required. Interestingly, the lack of a single N- or C-terminal MBT domain also resulted in a complete loss of repression. The lack of repression could not be attributed to differences in expression of the fusion proteins (Supplemental Fig. 2). Collectively, these findings indicate that all four MBT repeats of hSFMBT are required to induce potent transcriptional repression.

3.6. hSFMBT binding to histones H3 and H4 is mediated by the four MBT repeats

Recent structure studies suggest that the MBT domain closely resembles the chromodomains of HP1 and Pc [11,26], which selectively bind methylated histone H3 lysine 9 and 27, respectively [27–29]. Consistent with this, it was recently demonstrated that Drosophila SFMBT binds the mono- and dimethylated forms of histones H3 lysine 9 and H4 lysine 20 in vitro (5). To explore the possibility that human SFMBT could bind histones in vivo, the Gal4–DBD-4\cdot MBT or Gal4–DBD-null constructs were transfected into HEK-293 cells and immunoprecipitated with a Gal4–DBD antibody (Fig. 5A). Western analysis of the bound material indicates
Fig. 5. hSFMBT preferentially binds histones H3 and H4. (A) Western analysis was performed with a general histone H3, H4 or UBC9 antibody on 4% of the input, 5% and 10% of the Gal4–DBD immunoprecipitated material from HEK-293 cells transfected with a Gal4–DBD-null or Gal4–DBD-4x MBT plasmid. The data demonstrate that the four tandem MBT repeats of hSFMBT (4x MBT) specifically immunoprecipitated histones H3 > H4. (B) GST pull-down experiments with equal molar amounts of recombinant GST, GST–SFMBT or GST-4x MBT incubated with 50 μg of HEK-293 acid-extracted histones. Coomassie staining following SDS–PAGE demonstrates that the four MBT repeats of hSFMBT are sufficient for direct binding to histones H3 and H4. (C) GST pull-down experiments with recombinant GST-4x MBT and either myelin basic protein (MBP) or bovine serum albumin (BSA). GST-4x MBT failed to pull-down both MBP and BSA indicating a specific interaction between 4x MBT and histones H3 and H4.

3.7. hSFMBT binds the N-terminal tail of histone H3

To determine the region of histone H3 responsible for interacting with the MBT repeats of hSFMBT, GST pull-down experiments were performed with purified recombinant GST or GST-4x MBT and either acid-extracted histones or “tail-less” histones isolated from trypsinized HEK-293 oligonucleosomes [33]. Western analysis of the bound fractions using a general H3 antibody revealed that the four MBT repeats of hSFMBT failed to bind the H3 histone-fold region suggesting that the N-terminal tail is required for the interaction (Fig. 6A).

To confirm this hypothesis, co-immunoprecipitations were performed in HEK-293 cells using the full length FLAG-SFMBT construct and an N-terminal GST fusion construct containing the first 41 amino acids of human H3 (H3 1–41) [34]. While Western analysis of the FLAG-SFMBT immunoprecipitated material demonstrated that GST alone did not interact with hSFMBT, the GST–H3 1–41 fusion protein did bind hSFMBT (Fig. 6B). Importantly, a FLAG-null immunoprecipitation failed to bind GST–H3 1–41 and the GST–H3 1–41 fusion protein did not bind the FLAG beads indicating a specific interaction (Supplemental Fig. 3). To confirm these observations, the experiment was repeated using the same lysates but immunoprecipitated with glutathione-sepharose beads instead of FLAG beads. Western analysis of the GST–H3 1–41 immunoprecipitated material verified binding to FLAG-SFMBT (Fig. 6C). This was a specific interaction as the GST–H3 1–41 failed to immunoprecipitate UBC9. Collectively, these findings demonstrate that four MBT repeats of hSFMBT specifically bind the N-terminal tail of histone H3.
4. Discussion

Although Drosophila SFMBT was recently found to be a PcG protein required for Hox gene repression, the biological function of human SFMBT remained unknown [6]. In this study we demonstrate for the first time that hSFMBT is a nuclear matrix-associated protein that acts as a potent repressor of transcription. Our novel findings indicate that hSFMBT binds the N-terminal tail of histone H3 suggesting that this interaction is required for targeting hSFMBT to specific chromatin regions destined for repression. The tissue-restricted expression of hSFMBT to certain cell lineages further suggests that it functions to repress sets of genes required for specific developmental and differentiation programs, similar to other PcG proteins. Due to its functional similarity to dSfmbt, it is likely that hSFMBT also plays a role in PcG-mediated gene repression.

The amino acid sequence comparison between human and fly SFMBT revealed a high degree of conservation in the four tandem MBT repeats, suggesting that this region plays a critical role in protein function. Consistent with this hypothesis, we found that the four tandem MBT repeats of hSFMBT were necessary and sufficient for nuclear matrix-association, histone binding and transcriptional repression. Importantly, the lack of any one of the four MBT domains resulted in the abolishment of repressor activity indicating that all four repeats are required to form a functional structural unit that is necessary for biological activity. These findings are consistent with the crystal structure of hL(3)MBT where each of its three MBT repeats formed tight globular modules that interdigitate to create a novel three-leaved propeller-like structure [26]. Based on these findings, it is likely that the four tandem MBT repeats of hSFMBT also create a novel propeller-like structure that is required for functional activity.

One possible role of this structure is to interact with other members of a putative multi-protein complex. This is likely since most PcG family members are part of larger multimeric chromatin-associated complexes that are required for long term silencing of developmental and oncogenic genes [35,36]. Another possibility is that the MBT domains bind to specific histone modifications. Based on its homology to the histone methyllysine-binding Chromo and Tudor domains, it was previously proposed that the MBT domain could, likewise, bind modified histone tails [37]. The ligand binding pockets of the Chromo and Tudor domains bind the methylated lysine via a hydrophobic cage created by conserved aromatic residues within the motif [28,38]. Homologous aromatic residues are found within the MBT domain suggesting its ability to bind methylated histones. Indeed, it was recently reported that the MBT repeats of hL(3)MBT bound dimethylated H4 lysine 20 and, importantly, that the four tandem MBT repeats of dSfmbt preferentially bound the mono- and dimethylated lysine 278, 46968–46973. Therefore, it is possible that hSFMBT will also bind methylated histone tails; this will be determined in future experiments.

Acknowledgements: This work was supported by generous grants from the Robert E. and May R. Wright Foundation and the Donald E. & Delia B. Baxter Foundation. J.C.R. is a Pew Scholar in the Biomedical Sciences. R.C.T. is supported by National Institutes of Health (GM073839). The authors thank Michael Stallcup (University of Southern California) for the pGK1-luc, Gal4–DBD and CARM1 plasmids; Mitchell Lazar (University of Pennsylvania) for the pSV40-luc, pCMX-Gal4–DBD and SMRT plasmids; Woojin An (University of Southern California) for the GST–H3 1–41 plasmid and members of the Rice Lab for helpful suggestions and comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.06.025.

References


