AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-dependent transcription.
AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-dependent transcription

Hiroshi Okuda¹, Akinori Kanai², Shinji Ito³, Hirotaka Matsui²,† & Akihiko Yokoyama¹

Gene rearrangements generate MLL fusion genes, which can lead to aggressive leukemia. In most cases, MLL fuses with a gene encoding a component of the AEP (AF4 family/ENL family/P-TEFb) coactivator complex. MLL-AEP fusion proteins constitutively activate their target genes to immortalize haematopoietic progenitors. Here we show that AEP and MLL-AEP fusion proteins activate transcription through selectivity factor 1 (SL1), a core component of the pre-initiation complex (PIC) of RNA polymerase I (RNAP1). The pSER domain of AF4 family proteins associates with SL1 on chromatin and loads TATA-binding protein (TBP) onto the promoter to initiate RNA polymerase II (RNAP2)-dependent transcription. These results reveal a previously unknown transcription initiation mechanism involving AEP and a role for SL1 as a TBP-loading factor in RNAP2-dependent gene activation.

¹Laboratory for Malignancy Control Research, Kyoto University Graduate School of Medicine, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8501, Japan. ²Department of Molecular Oncology and Leukemia Program Project, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan. ³Medical Research Support Center, Kyoto University Graduate School of Medicine, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. †Present address: Department of Molecular Laboratory Medicine, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Chuo-ku, Kumamoto 860-8556, Japan. Correspondence and requests for materials should be addressed to A.Y. (email: yokoyama@dsk.med.kyoto-u.ac.jp).
Chromosomal translocations generate a variety of MLL (also known as KMT2A, MLL1, HRX and ALL-1) fusion genes, which cause acute leukemia in myeloid and lymphoid lineages. Although >70 different fusion partners have been identified, the majority of leukemia cases are caused by the chimeric genes formed by MLL and a gene encoding a component of the AEP (AF4 family/ENL family/P-TEFb) coactivator complex. The AEP complex comprises AF4 family proteins (for example, AF4 and AF5Q31), ENL family proteins (for example, ENL and AP9) and the P-TEFb elongation factor. Similar, if not identical, complexes have been identified and shown to play important roles in various biological processes (for example, heat shock response and transcription of the HIV viral genome). AEP associates with RNA polymerase II (RNAP2)-specific factors, including the polymerase II-associated factor 1 complex and the mediator complex, and thus appears to be closely linked to RNAP2-dependent transcription. MLL–AEP fusion proteins constitutively activate their target genes by recruiting AEP components to the target chromatin, whereas wild-type MLL recruits AEP in a context-dependent manner. In the haematopoietic lineage, MLL fusion proteins aberrantly activate a subset of genes implicated in the haematopoietic stem cell programme, such as HOXA9 and MEIS1 (ref. 9). Constitutive expression of these genes in haematopoietic progenitors has been shown to induce leukemia in a mouse model, suggesting that a gain-of-function mechanism underlies the development of MLL leukemia. MLL fusion proteins form a complex with MENIN and LEDGF, and the MLL fusion protein complex directly binds to promoter-proximal pausing. However, it remains largely unknown how MLL–AEP fusion proteins activate their target genes.

Here we report that a serine-rich domain in AF4 family proteins, termed pSER, is an essential functional component of MLL–AEP fusion-dependent gene activation and leukemic transformation. Through biochemical purification, we identified selectivity factor 1 (SL1) as a novel factor associated with the pSER domain. SL1, comprising TATA-binding protein (TBP) and four TATA box-binding protein-associated factors (TAFI41), is a core component of the pre-initiation complex (PIC) of RNA polymerase I (RNAP1; refs 18–21). In the presence of upstream binding factor (UBF), SL1 forms a PIC on the promoters of ribosomal RNA genes to drive RNAP1-dependent transcription. However, it is unknown whether SL1 plays a role in RNAP2-dependent transcription. Our results indicate that the AEP coactivator complex facilitates the initiation of RNAP2-dependent transcription via SL1 activity by loading TBP onto the TATA element. MLL–AEP fusion proteins use this TBP-loading function to activate transcription in leukemic transformation, whereas the wild-type AEP complex activates gene expression in the same manner under physiological conditions.

Results

The pSER domain drives myeloid transformation. In ex vivo conditions, MLL fusion proteins transform myeloid progenitors by constitutively activating haematopoietic stem cell programme genes such as Hoxa9 (ref. 23). As transformation leads to the immortalization of myeloid progenitors, it is a critical event in leukemogenesis induced by MLL fusion proteins. Their transforming properties are evidenced by sustained expression of Hoxa9 in the first round colonies and vigorous colony-forming activities in the third and fourth rounds of replating in myeloid progenitor transformation assays (Fig. 1a,b). The minimal functional domains of the fusion partner portions of MLL–ENL and MLL–AF5Q31 required for transformation are the ANC1 homology domain (AHD) of ENL (also known as MLLT1) and the carboxy-terminal homology domain (CHD) of AF5Q31 (also known as AFF4), respectively (Fig. 1b). Immunoprecipitation (IP) followed by western blotting (WB) showed that both of the fusion partner portions serve as a binding platform for the fusion protein (also known as AFF1). The ENL portion also associated with DOT1L, another ENL-associated factor implicated in MLL fusion-dependent leukemogenesis.

On the other hand, MLL–AF5Q31–4 did not pull down DOT1L, indicating that direct recruitment of DOT1L is not critical for transformation (Fig. 1c). Removal of the AF4-binding platform resulted in loss of transforming activity (Fig. 1b). Thus, recruitment of AF4 appears to be essential for MLL–AEP fusion-dependent gene activation and transformation. With that in mind, we inferred that one or more functional domains in AF4, besides the CHD, are responsible for MLL–AEP fusion-dependent gene activation and transformation. Recently, we identified the minimum targeting module of MLL–AEP fusion proteins that is sufficient for the recognition of target chromatin. The module comprises the PWWP domain and the CXXC domain (Fig. 1b). An artificial gene in which the minimum targeting module was fused to the AHD of ENL or to the CHD of AF5Q31 activated Hoxa9 and immortalized myeloid progenitors (Fig. 1b). To investigate the role of the functional domains of AF4 in MLL fusion-dependent transformation, we constructed a series of mutants in which the minimum targeting module was tethered to the subdivided domains of AF4, termed AF4-1, AF4-2N, AF4-2C and AF4-3 (Fig. 1b,d), and examined their transforming ability. Of the four AF4 domains, only the AF4-2C domain, which encompasses the evolutionarily conserved pSER domain, exhibited transforming abilities (Fig. 1b). These results suggest that the pSER domain of AF4 mediates the transformation of myeloid progenitors induced by MLL–AEP fusion proteins.

The pSER domain associates with SL1 on chromatin. Next, we characterized the transcriptional properties of each AF4 domain. To this end, we generated a series of constructs in which a GAL4 DNA-binding domain was fused to each subdivided domain of AF4. Transactivation assays using a GAL4-responsive reporter showed that the FLAG-tagged GAL4–AF4–2C fusion protein (gAF4–2C) had substantial transactivation activity (Fig. 2a), consistent with previous reports. As transcription takes place on chromatin in vivo, we inferred that the pSER domain associated with cofactors on chromatin to activate transcription. To purify the chromatin protein complex, we used the fractionation-assisted native chromatin IP (fanChIP) method, which we previously established. In this method, chromatin-unbound materials are removed by cytoskeleton buffer extraction, and chromatin and chromatin-bound materials are solubilized by micrococcal nuclease (MNase) digestion followed by exposure to detergent (Supplementary Fig. 1a). The factors that associated specifically with gAF4–2C on chromatin were copurified from the chromatin fraction by affinity purification using an anti-FLAG antibody (Supplementary Fig. 1b). Mass spectrometry analysis of the purified materials showed that all components of SL1 bound to the pSER domain (Fig. 2b). FanChIP analysis
followed by WB (fanChIP–WB) confirmed that the pSER domain associated specifically with endogenous TAF1C and TBP, but not with TFIIH, TAF1 (TAFp1p250), RRN3 or RAP1 (POLR1A; Supplementary Fig. 1c), whereas the other domains exhibited their own cofactor binding properties (for example, AF4-1-bound CDK9 (a component of P-TEFb) and AF4-2N-bound Eleven-nineteen lysine-rich leukemia protein (ELL); Fig. 2c). The pSER domain pulled down all the components of SL1 that were exogenously expressed, either individually (Supplementary Fig. 1d) or simultaneously (Fig. 2d). This interaction occurred only on chromatin (Fig. 2e) but was not dependent on DNA (Fig. 2f). Wild-type AF4 also pulled down TAF1C, along with ENL and CDK9 (Fig. 2g). Moreover, MLL–ENL and MLL–AF5Q31-4 pulled down TAF1C when co-expressed with MENIN and AF4 (Fig. 2h). Hence, both AEP and the MLL–AEP fusion proteins specifically associate with SL1 on chromatin.
Figure 2 | The pSER domain associates with SL1 on chromatin and activates transcription. (a) Transactivation activities of the AF4 subdomains. The schematic structures of various FLAG-tagged GAL4 fusion proteins are shown. Various GAL4 fusion genes were cloned into pCMV5 vector and used for transient expression. fG, FLAG-tagged GAL4 DNA-binding domain. 293T cells were transiently transfected with the expression vectors for various GAL4 fusion proteins with the pFR-LUC and pRL-TK reporter plasmids. Promoter activity was assessed with the dual luciferase reporter assay. The transcription activation activity, normalized to the RL activity, is shown with error bars (s.d. from triplicates) relative to the value of fG-AF4-2C (arbitrarily set at 1). UAS, upstream activation sequence. (b) The protein scores of SL1 components identified by mass spectrometry as co-precipitates with fG–AF4-2C. (c) Factors associated with various AF4 subdomains. Various FLAG-tagged GAL4–AF4 fusion proteins were expressed in 293T cells and subjected to fanChIP–WB. The sample shown in the input lane is indicated by an asterisk. (d) Association of the SL1 complex with the pSER domain. All SL1 components were co-expressed with fG-AF4-2C in 293T cells and subjected to fanChIP–WB. The pCMV5 HisMax-TAF1A/B/D and -TBP vectors and the pCMV5-TAF1C-HA vector were used for transient expression of Xpress-tagged or HA-tagged SL1 components. (e) Chromatin-specific association of SL1 with the pSER domain. IP–WB analyses were performed with the chromatin fraction (CHR) or the chromatin-unbound soluble faction (SOL) from 293T cells expressing fG-AF4-2C. (f) DNA-independent association of SL1 with the pSER domain. FanChIP analysis, as described in c, was performed, followed by DNASEI treatment. (g) Association of SL1 with AEP, FLAG-tagged AF4 was transiently expressed in 293T cells and subjected to fanChIP–WB. The pBICEP2-AF4 vector was used for transient expression. (h) Association of SL1 with MLL–AEP fusion proteins. FLAG-tagged MLL–AEP fusion proteins were co-expressed with Xpress-tagged AF4 (xAF4) and HA-tagged MENIN (MENINh) in 293T cells and subjected to fanChIP–WB. The pCMV5–MLL fusion vectors, the pcDNA 3.1 hygro (+)-MEN1-HA vector and the pcDNA4 HisMax-AF4 vector were used for transient expression.

MLL–ENL co-localizes with TAF1C at the target promoters. To examine whether MLL–AEP fusion proteins co-localize with SL1 at target promoters, we performed a fanChIP assay followed by deep sequencing (fanChIP-seq) with HB1119 cells, which harbour a t(11;19) translocation and therefore express the MLL–ENL protein. From the enriched ChIP signals obtained using an anti-MLL antibody (Fig. 3a and Supplementary Fig. 2a), we identified a representative set of MLL target genes (112 genes), which included many previously identified direct MLL target genes such as HOXA9, MEIS1, RUNX2, CDKN1B and CDKN2C3,33 (Supplementary Table 1). The average distribution of the MLL proteins at the MLL target promoters suggested that MLL–ENL associated with chromatin near the transcription start sites (Fig. 3b). ChIP signals for AF4 and TAF1C were also observed at the MLL target promoters (Fig. 3a,b and Supplementary Fig. 2a). The signal intensity of TAF1C was highly correlated with those of MLL and AF4 at the promoter proximal regions of the MLL target loci (Supplementary Fig. 2b). This was confirmed by the results of a fanChIP assay followed by quantitative PCR (qPCR) analysis (fanChIP–qPCR), performed for several MLL target loci (Fig. 3c). FanChIP–qPCR analysis using anti-MENIN, anti-CDK9 and anti-TBP antibodies demonstrated that the MLL–ENL/AF4/SL1 complex formed at the MLL target promoters. RNA2p and its cofactors, such as NELF, also localized at the MLL target promoters (Fig. 3c). Expression of these genes was sensitive to α-amanitin treatment (Supplementary Fig. 2c), suggesting that
RNAP2-dependent transcription occurs at MLL–ENL target genes. These results indicate that SL1 facilitates, rather than inhibits, RNAP2-dependent transcription at MLL target genes (Fig. 3d).

**TAF1C is required for AEP-dependent gene activation.** To examine whether SL1 is required for MLL–ENL-dependent gene activation, we knocked down Taf1c in MLL–ENL-transformed cells by using two different short hairpin RNAs (shRNAs). Knockdown of Taf1c decreased the expression of MLL–ENL target genes (Hoxa9, Hoxa10 and Runx2; Fig. 4a). These results suggest that TAF1C is an essential cofactor for MLL–AEP fusion-dependent gene activation.

Next, we tested whether SL1 is required for AEP-dependent transcription in a non-oncogenic context. Taf1c was knocked down in immortalized mouse embryonic fibroblasts (IMEFs). As wild-type MLL collaborates with AEP at many target genes, the expressions of Hoxc8, Hoxc9, Cdkn1b and Cdkn2c are dependent...
Figure 4 | TAF1C is required for gene activation by MLL-ENL and AEP. (a) The expression of MLL-ENL target genes (Hoxa9, Hoxa10 and Runx2) after knockdown of Taf1c with two different shRNAs in MLL-ENL-transformed cells. The expression level normalized to Tbp is shown relative to the value of the vector control (arbitrarily set at 100%) with error bars (s.d. of PCR triplicates). Puro, puromycin. (b) Effect of Taf1c knockdown on AEP-dependent gene activation. Taf1c was knocked down with two different shRNAs in iMEFs. The expression of AEP target genes (Hoxa8, Hoxc9, Cdkn1b and Cdkn2c) was analysed with reverse transcriptase (RT)-qPCR as described in a. (c) Gene set enrichment analysis (GSEA) of the expression profiles of iMEFs after knockdown of Taf1c (#1 and #2), Enl and Mll. Genes that exhibited a greater than threefold decrease on knockdown in RNA sequencing (RNA-seq) were defined as target genes (Supplementary Fig. 3c,d). The target gene sets of ENL, MLL and TAF1C (shRNA#1) were downregulated by knockdown of Taf1c with shRNA#2. Expression levels of the target genes were also shown by scatter plots. Target genes of ENL, MLL and TAF1C are highlighted in blue. FDR, false discovery rate; NES, normalized enrichment score; Nom-P-value, nominal P-value; RPKM, reads per kilobase of exon per million mapped reads.

on both ENL and MLL in iMEFs (Supplementary Fig. 3a,b). shRNAs against Taf1c decreased the expression of the MLL/AEP target genes (Fig. 4b). Moreover, messenger RNA sequencing analysis showed that knockdown of Taf1c globally downregulated ENL and MLL target genes in iMEFs (Fig. 4c and Supplementary Fig. 3c–g). These results indicate that SL1 is required for gene activation mediated by the MLL/AEP axis.

Unique mode of transcription initiation by the pSER domain. To examine whether the initiation of RNAP2-dependent transcription by the pSER domain is responsible for MLL–AEP fusion-dependent transformation, we generated artificial genes in which the minimum targeting module of MLL was fused to various transactivation domains (ADs) that have been shown to initiate RNAP2-dependent transcription. We tested their
transferring ability in myeloid progenitor transforming assays (Fig. 5a,b). The AD of VP16 recruits mediators, the basic transcriptional machinery and CBP/p300 coactivators to initiate RNAP2-dependent transcription. The AD of MLL also recruits mediators, the basic transcriptional machinery and CBP/p300 coactivators to initiate RNAP2-dependent transcription 34–37. The AD of MLL also showed that only the pSER domain formed a complex with SL1 (Fig. 5d). To analyse transcription activation in a chromatin context, we established a 293T cell line in which the GAL4-responsive reporter cassette was inserted into the genome. Transactivation assays using this cell line showed that the pSER domain retained its transactivation activity in a chromatin context, which was the weakest of the three ADs (Fig. 5c) and was susceptible to α-amanitin treatment (Supplementary Fig. 4b). FanChIP–qPCR analysis showed that both TAF1C and TBP were efficiently recruited to the promoter through the pSER domain, whereas the ADs of MLL and VP16 recruited TBP, but not TAF1C, to the promoter (Fig. 5e). TAF1, a component of the TFIID complex80, was recruited to the promoter by the ADs of

Figure 5 | Transcription initiation by the pSER domain confers transforming properties. (a) Transforming abilities of P’C’ + mutants fused with various ADs. Hoxa9 expression normalized to Gapdh is shown relative to the value of P’C’ + h-AF4-2C (arbitrarily set at 100%) with error bars (s.d. of PCR triplicates). The colony-forming abilities were analysed as described for Fig. 1b. (b) Protein expression of the P’C’ + mutants in the virus packaging cells. (c) Transcription activation activity of various GAL4–AD fusion proteins on the chromatinized GAL4-responsive promoter. 293T–LUC cells were transiently transfected with the pCMV5 expression vectors for various FLAG-tagged GAL4 fusion proteins and the pRL–TK reporter plasmid. Promoter activity was assessed with the dual luciferase reporter assay. The normalized transcription activation activity is shown in the logarithmic scale relative to the value of fG (arbitrarily set at 1). (d) Association of SL1 components with various transcriptional activation domains. FLAG-tagged GAL4–AD fusion proteins were expressed in 293T–LUC cells and subjected to fanChIP–WB. The sample shown in the input lane is indicated by an asterisk. (e) Localization of FLAG-tagged GAL4–AD fusion proteins, TAF1C, TBP and TAF1 on the chromatinized luciferase promoter. Each FLAG-tagged GAL4–AD fusion protein was expressed in 293T–LUC cells carrying the FR–LUC reporter cassette inserted in the genome and subjected to fanChIP–qPCR with the indicated antibodies. Precipitated DNA was analysed with qPCR, using probes specific to the pre-transcription start site (TSS), TSS and post-TSS regions. (f) Models of alternative RNAP2-dependent transcription initiation by SL1 and TFIID.

The colony-forming abilities were analysed as described for Fig. 1b. (c) Transcription activation activity of various GAL4–AD fusion proteins on the chromatinized GAL4-responsive promoter. 293T–LUC cells were transiently transfected with the pCMV5 expression vectors for various FLAG-tagged GAL4 fusion proteins and the pRL–TK reporter plasmid. Promoter activity was assessed with the dual luciferase reporter assay. The normalized transcription activation activity is shown in the logarithmic scale relative to the value of fG (arbitrarily set at 1). (d) Association of SL1 components with various transcriptional activation domains. FLAG-tagged GAL4–AD fusion proteins were expressed in 293T–LUC cells and subjected to fanChIP–WB. The sample shown in the input lane is indicated by an asterisk. (e) Localization of FLAG-tagged GAL4–AD fusion proteins, TAF1C, TBP and TAF1 on the chromatinized luciferase promoter. Each FLAG-tagged GAL4–AD fusion protein was expressed in 293T–LUC cells carrying the FR–LUC reporter cassette inserted in the genome and subjected to fanChIP–qPCR with the indicated antibodies. Precipitated DNA was analysed with qPCR, using probes specific to the pre-transcription start site (TSS), TSS and post-TSS regions. (f) Models of alternative RNAP2-dependent transcription initiation by SL1 and TFIID.
MLL and VP16 but not by the pSER domain (Fig. 5e). These results suggest that the pSER domain uses SL1 exclusively as the primary TBP recruiting factor to activate transcription initiation, whereas the ADs of MLL and VP16 use TFIID for TBP recruitment (Fig. 5f).

The pSER domain loads TBP onto the TATA element via SL1. To evaluate the importance of the TATA element in pSER-dependent transactivation, we generated a reporter plasmid lacking the TATA element (Fig. 6a). The TATA-deleted reporter (dTATA) or the TATA-containing reporter (control) was stably inserted into the genome of 293T cells and the transactivation activities of fG–AF4-2C were measured. The pSER-dependent transactivation activity on the dTATA reporter was drastically impaired, compared with that on the control reporter (Fig. 6b). Similar results were obtained when reporter plasmids with or without the TATA element were transfected into 293T cells (Supplementary Fig. 5), arguing against the possibility that the observed decrease in transcription from the dTATA reporter was due to the different genomic positions of the reporter cassettes. The ADs of MLL and VP16 behaved in a similar manner (Fig. 6b and Supplementary Fig. 5), suggesting that all three transactivation modules require the presence of the TATA element for efficient transcription in this promoter context. The localization of the GAL4 fusion proteins to the GAL4-responsive elements was not affected by the absence of a TATA element (Fig. 6c). Regardless of the presence of a TATA element, the pSER domain recruited TAF1C and TBP to the promoter. On the other hand, TBP recruitment was severely impaired in MLL–AD- or VP16–AD-dependent transactivation on the dTATA reporter, suggesting that TBP was not stably tethered to the dTATA promoter as part of the TFIID complex. These results indicate that the pSER domain tethers SL1 on chromatin in a TATA element-independent manner and loads TBP onto the TATA element to initiate transcription in this promoter context.

**Figure 6 | The TATA element potentiates pSER-dependent transcriptional activation.** (a) Schematic structures of the GAL4-responsive reporter with or without the TATA element. The FR-LUC reporter is termed ‘Control’ and the TATA-deficient version lacking the TATA sequence of the TATA element is termed ‘dTATA’. (b) Requirement for the TATA element for transactivation by various ADs. 293T–RL–LUC cells carrying the TK–RL reporter and the FR–LUC reporter, with (control) or without (dTATA) the TATA element, were transiently transfected with the pCMV5 expression vectors for various GAL4 fusion proteins. Promoter activity was assessed with the dual luciferase reporter assay. Transcription activation activities, normalized to the RL activity, are shown relative to the value of each control (arbitrarily set as 100). (c) Localization of FLAG-tagged GAL4–AD fusion proteins, TAF1C and TBP on the chromatinized luciferase promoter with or without the TATA element. Each FLAG-tagged GAL4–AD fusion protein was expressed in 293T–RL–LUC (control or dTATA) cells and subjected to FanChIP–qPCR using the indicated antibodies. Precipitated DNA was analysed with qPCR, using a probe specific to the pre-transcription start site (TSS) region, as shown in a.

**Discussion**

In this study, we identified an additional important layer of complexity in the effects of MLL fusion proteins on target genes. Our results indicate that the recruitment of AF4 family proteins is a critical first step in MLL–AEP fusion-dependent leukemic transformation. The pSER domain of AF4 associated with SL1 on chromatin to activate RNA Pol2-dependent transcription by loading TBP onto the promoter. The results show that SL1 is a TBP-loading factor involved in gene activation induced by MLL–AEP fusion proteins and wild-type AEP.

AEP contains the P-TEFb elongation factor and associates with ELL family proteins, both of which exhibit transcription elongation activity. Therefore, the AEP complex is also referred to as the super elongation complex. These discoveries led to the suggestion that transcription elongation activity plays an essential role in MLL–AEP fusion-dependent leukemic transformation. However, our detailed structure/function analyses showed that neither of the binding platforms for P-TEFb or ELL conferred transforming abilities (Fig. 1b and 2015 Macmillan Publishers Limited. All rights reserved.
Figure 7 | The SDE and NKW motifs collaborate to initiate transcription. (a) Alignment of the amino acid sequences of the SDE and NKW motifs of human AF4 family proteins. Multiple alignments were performed using the Clustal W software. ‘’, 100% amino acid identity; ’.’, 75% amino acid similarity. The SDE and NKW motifs are shown in open red boxes. (b) Transforming ability of P’C’+ mutants fused with various pSER subdomains. Hoxa9 expression in the first round colonies and the colony-forming units (CFUs) at the third and fourth rounds of replating are shown as described in Fig. 5a. (c) Association of SL1 components with various pSER subdomains. FLAG-tagged GAL4-pSER subdomains fusion proteins were expressed in 293T–LUC cells and subjected to fanChIP–qPCR as described in Fig. 5e. (d) Transactivation assay was performed as described in Fig. 5c. (e) Association of SL1 components with various pSER subdomains. FLAG-tagged GAL4-pSER subdomains fusion proteins were expressed in 293T–LUC cells and subjected to fanChIP–qPCR as described in Fig. 5e.

Fig. 2c), indicating that transcription elongation activity is not sufficient for MLL–AEP fusion-dependent gene activation. Instead, the pSER domain, which harbours transcription initiation activities, conferred transforming abilities to MLL–AEP fusion proteins. Hence, the AEP complex is more than an elongation complex: it serves as a multi-functional transcriptional coactivator that facilitates many steps of the transcription cycle.

DOT1L has been shown to play important roles in MLL fusion-dependent leukemic transformation13. DOT1L recruitment through direct interaction enhances transforming potentials14. Consistent with these reports, the clonogenicities of the AF5Q31 CHD fusion proteins, which is deficient for DOT1L recruitment, were weaker than those of the ENL AHD fusion proteins (Fig. 1b,c). These results suggest a supporting role of DOT1L for AEP-dependent gene activation, which is probably mediated by its histone methyltransferase activity that establishes the chromatin environment repelling to SIRT1-dependent gene silencing15.

SL1, which comprises TBP and four TAF1 proteins, confers promoter selectivity to RNAPI (ref. 22). On the other hand, the TFIID complex, which comprises a number of TAF1 proteins, plays a major role in RNAPII-dependent transcription16. As TBP associates with TAF1S and TAF1H in a mutually exclusive manner19, it has long been thought that the functions of SL1 were restricted to RNAPI1-dependent transcription. However, our biochemical analyses identified SL1 as a specific binding partner of the AEP coactivator complex that activates...
observations and assumptions, we infer that (1) AF4 family proteins recruit SL1 to the target chromatin primarily through the SDE motif, (2) further association through the NKW motif changes the conformation of chromatin-bound SL1, releasing TAF1B, (3) and TAF1B-deficient SL1 is bound by TFIIH, which forms the PIC of RNAP2 to initiate transcription (Fig. 8). The presence of a TATA element that conforms to the ~90° bent structure required for the PIC formation potentiates this process (Fig. 6). Thus, SL1 might facilitate probing for DNA sequences suited to TBP-induced bending. SL1 might induce a similar bent structure to form the PIC of RNAP1 (ref. 51) at the ribosomal RNA promoters that lack a TATA element. Most gene promoters do not contain TATA elements that satisfy the TATA box consensus sequence. It appears that TATA-like AT-rich sequences serve as alternative TBP-binding sequences to promote transcription at TATA-less promoters. The AEP/SL1 complex might facilitate the formation of the TBP–DNA complex and the subsequent induction of DNA bending not only at the TATA element but also at TATA-like AT-rich sequences.

In summary, our study identified a novel cofactor of MLL–AEP fusion proteins and the AEP coactivator complex. To our surprise, the essential cofactor required for AEP-dependent gene activation is SL1, a component of the PIC of RNAP1. These results not only provide significant insights into the molecular mechanisms underlying MLL fusion-mediated leukemogenesis but also shed light on the previously unrecognized involvement of SL1 in RNAP2-dependent transcription.

Figure 8 | A working model. A working model of AEP-dependent transcriptional activation through SL1.

RNAP2-dependent transcription (Fig. 2). The MLL/AEP axis maintains the expression of cellular memory genes, such as Hox genes (Fig. 4b,c) and Men1 genes in iMEFs, cells that make up the hematopoietic stem cell pool, and express the AEP oncogene. The NKW motif is essential for the activation of RNAP2-dependent transcription. One is the SDE motif, the primary binding platform for SL1. The amino acid sequence of this motif is similar to that of the SL1-binding motif in UBF. Therefore, AF4–SL1 binding probably occurs through a mechanism similar to that underlying UBF–SL1 binding. The other critical motif is the NKW motif. Despite our extensive efforts to identify proteins that specifically bind this motif, none has yet been identified. Although we do not rule out the possibilities that other transcriptional regulators are recruited to the NKW motif, we infer that the NKW motif binds to the surface of SL1 itself to induce conformational changes. TAF1B is structurally similar to TFIIB and is predicted to position itself in the RNAP1–PIC, similar to TFIIA in the RNAP2–PIC. Therefore, for RNAP2-dependent transcription to proceed following the binding of SL1 to the promoter, TFIIA must replace TAF1B as the binding partner of TBP. Given these observations and assumptions, we infer that (1) AF4 family proteins recruit SL1 to the target chromatin primarily through the SDE motif, (2) further association through the NKW motif changes the conformation of chromatin-bound SL1, releasing TAF1B, (3) and TAF1B-deficient SL1 is bound by TFIIH, which forms the PIC of RNAP2 to initiate transcription (Fig. 8). The presence of a TATA element that conforms to the ~90° bent structure required for the PIC formation potentiates this process (Fig. 6). Thus, SL1 might facilitate probing for DNA sequences suited to TBP-induced bending. SL1 might induce a similar bent structure to form the PIC of RNAP1 (ref. 51) at the ribosomal RNA promoters that lack a TATA element. Most gene promoters do not contain TATA elements that satisfy the TATA box consensus sequence. It appears that TATA-like AT-rich sequences serve as alternative TBP-binding sequences to promote transcription at TATA-less promoters. The AEP/SL1 complex might facilitate the formation of the TBP–DNA complex and the subsequent induction of DNA bending not only at the TATA element but also at TATA-like AT-rich sequences.

Methods Vector construction. The pMSCV-neo-FLAG-MLL-ENL vector and its derivative were generated through restriction enzyme digestion/PCR-based mutagenesis. The complementary DNAs were cloned into the pMSCV neo vector (for virus production; Clontech) or into the pcMV5 vector and the pcDNA4 HisMax vector (for transient expression). The pFR-LUC plasmid with the deletion of the TATA element was made by GeneArt Gene Synthesis (Life Technologies) from the pFR-LUC plasmid (Agilent). The plKO1-puro-FR-LUC vectors and its blasticidin version, with or without the TATA element, were generated by restriction enzyme digestion/PCR-based mutagenesis of the pRL–TK vector and the pcDNA HisMax vector (for transient expression). The pfr-LUC plasmid with the deletion of the TATA box was made by GeneArt Gene Synthesis (Life Technologies) from the pFR-LUC plasmid (Agilent). The plKO1-puro-FR-LUC vectors and its blasticidin version, with or without the TATA element, were generated by restriction enzyme digestion/PCR-based mutagenesis of the pRL–TK plasmid (Promega), the plKO1 vector (GE Healthcare) and a blasticidin cassette. pLKO.1-hygro–TK–RL vector and its zeocin version, which contains the herpes simplex virus thymidine kinase promoter, the Renilla luciferase (RL) coding sequence and a hygromycin or zeocin cassette was generated by restriction enzyme digestion/PCR-based mutagenesis of the pRL–TK plasmid (Promega), the plKO1 vector and a hygromycin or zeocin cassette. The shRNA expression vectors, targeting murine Enl (Enl (TRCN0000084405), Men1 (TRCN0000034394), Mll (TRCN0000034426), Taf1c1 (TRCN0000082215) and Taf1c2 (TRCN0000082217) were obtained from GE Healthcare.

Cells and cell culture. The human leukemia cell line HB1119 (ref. 3) was cultured in RPMI 1640 medium, supplemented with 10% fetal bovine serum (FBS) and penicillin–streptomycin (PS). The 293T and 293TN (System Biosciences) cell lines and iMEFs were cultured in DMEM medium supplemented with 10% FBS and PS. To exclude the effects of the mosaic expression of Hox genes in iMEFs, cells were cloned once to establish a cell line that expressed a certain set of Hox genes homogeneously. Ecotropic virus packaging cells (PLAT-E cells) were cultured in DMEM supplemented with 10% FBS, puromycin, blasticidin and PS. 293T–LUC cells were generated by co-transfection of the lentivirus carrying plKO-LUC vector and its blasticidin version, with or without the TATA element. 293T–LUC cells were generated by sequential transduction of the plKO-hygro–TK–RL reporter and the plKO-puro-FR-LUC reporter with or without the TATA element. 293T–LUC–G reporter cells were generated by sequential transduction of the plKO-zeo–TK–RL reporter, the plKO-bla-FR-LUC reporter and the pMSCV-hygro–FLAG4 fusion vectors.

Antibodies. The antibodies used for the assays are described in Supplementary Table 2. The uncropped WBs are provided in Supplementary Figs 6–11.

Virus production. The ecotropic retrovirus was produced using PLAT-E packaging cells. The lentivirus was produced in 293TN cells using the pMD2.g/pRRE, pRSV-rev and pMD2.G vectors. The supernatant medium containing the virus was harvested 24–48 h following transduction and used in viral transduction.
**Myeloid progenitor transformation assay.** The myeloid progenitor transformation assay was performed using cells harvested from the femurs and tibiae of 5-week-old female C57BL/6 mice.**1** C-Kit-positive cells were washed five times with washing buffer (1:1 mixture of EDTA buffer and MNase buffer with 20 mM EDTA) and then eluted in elution buffer (1% SDS and 50 mM NaHCO3). The eluted materials were analysed with various methods, including WB, qPCR, deep sequencing and mass spectrometry. Deep sequencing of the precipitated DNA was performed using a TrueSeq Chip Sample Prep Kit (Illumina) and Genome Analyzer IIx (Illumina) at the core facility of Hiroshima University. The data were analysed using the Integrative Genome Viewer (Broad Institute). qPCR analysis of the precipitated DNA was performed using the custom-made primer sets described in Supplementary Table 4. The value relative to the control was determined using a standard curve and the relative quantification method. Optionally, the precipitates were equilibrated with MNase buffer, treated with DNase I (Qiagen) for 10 min at 37 °C and washed five times with washing buffer to remove the DNA in the sample. The precipitates were analysed with WB or SYBR green staining.

**mRNA sequencing.** Total RNA was prepared using the RNeasy kit (Qiagen) and reverse transcribed using a Superscript III First Strand cDNA Synthesis Kit, with oligo(dT) primers or random hexamers (for experiment of α-aminant treatment; Life Technologies). Gene expression was confirmed with qPCR, using the TaqMan probes described in Supplementary Table 3 (Life Technologies). The expression levels, normalized to those of Gapdh, Tbp or 18 S rRNA, were determined using a standard curve and the relative quantification method, as described in ABI User Bulletin #2.

**Immunoprecipitation.** The expression vectors for MLL fusion proteins and the cofactors were transfected into 293T cells using the Lipofectamine 2000 reagent (Life Technologies). The cells cultured in a 10-cm dish were suspended in 1 ml of buffer (50 mM Tris-HCl pH 7.5, 4 mM MgCl2, 1 mM CaCl2, 0.3 M sucrose, 5 mM sodium butyrate, 0.5 mM DTT and a protease inhibitor cocktail) and treated with MNase. The MNase reaction was stopped by adding EDTA (pH 8.0) at a final concentration of 20 mM. Lysis buffer was then added to increase solubility. The precipitates were analysed with WB using a SpeedVac (Thermo) for 30 min at 50 °C.

**Liquid chromatography–tandem mass spectrometry analysis.** Purified proteins were visualized with Oriole staining (Bio-Rad) after SDS–PAGE analysis. The pieces of gel containing precipitating proteins were cut out and washed with 50 mM of ammonium hydrogen carbonate containing 50% acetonitrile. The gel pieces were dried using a SpeedVac (Thermo) and suspended in 50 mM of ammonium hydrogen carbonate. The proteins were deoxidized and acetylated with the addition of 10 mM DTT and 50 mM iodoacetamide, and then digested with trypsin at 37 °C for 18 h in a rotating chamber. The beads were washed five times with 500 μl of lysis buffer. The precipitated proteins were harvested in elution buffer (1% SDS, 50 mM NaHCO3). The eluted samples were mixed with an equal amount of 2 x SDS–PAGE sample buffer and then subjected to WB.

**Statistical analysis.** Correlation coefficient was calculated by Pearson’s correlation analysis using GraphPad Prism 6.0 (GraphPad Software).

**References**


5. Sobbhan, B. et al. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. **Mol. Cell** 38, 308–322 (2010).


