Cell, Vol. 115, 293–303, October 31, 2003, Copyright ©2003 by Cell Press

# Taspase1: A Threonine Aspartase Required for Cleavage of MLL and Proper *HOX* Gene Expression

**James J.-D. Hsieh, Emily H.-Y. Cheng, and Stanley J. Korsmeyer***
Howard Hughes Medical Institute
Dana-Farber Cancer Institute
Harvard Medical School
Boston, Massachusetts 02115

## Summary

**The *Mixed-Lineage Leukemia* gene (*MLL*/*HRX*/*ALL1*) encodes a large nuclear protein homologous to *Drosophila trithorax* that is required for the maintenance of *HOX* gene expression. MLL is cleaved at two conserved sites generating N320 and C180 fragments, which heterodimerize to stabilize the complex and confer its subnuclear destination. Here, we purify and clone the protease responsible for cleaving MLL. We entitle it Taspase1 as it initiates a class of endopeptidases that utilize an N-terminal threonine as the active site nucleophile to proteolyze polypeptide substrates following aspartate. Taspase1 proenzyme is intramolecularly proteolyzed generating an active 28 kDa α/22 kDa β heterodimer. RNAi-mediated knockdown of Taspase1 results in the appearance of unprocessed MLL and the loss of proper *HOX* gene expression. *Taspase1* coevolved with *MLL*/*trithorax* as *Arthropoda* and *Chordata* emerged from *Metazoa* suggesting that Taspase1 originated to regulate complex segmental body plans in higher organisms.**

## Introduction

*MLL*/*HRX*/*ALL1* encodes a 3969 amino acid nuclear protein bearing multiple conserved domains with assigned activities including an N terminus with three AT-hook motifs that bind AT rich DNA segments (Zeleznik-Le et al., 1994), a DNA methyl transferase homology domain that represses transcription (Xia et al., 2003), four PHD fingers that mediate protein-protein interactions (Fair et al., 2001), a transactivation domain that interacts with CBP (Ernst et al., 2001), and a C-terminal SET domain with histone H3 lysine 4 methyl transferase activity (Milne et al., 2002; Nakamura et al., 2002) (Figure 1A). *MLL* and its *Drosophila* homolog *trithorax* are required for maintaining proper *Hox* and *homeotic* gene expression patterns, respectively (Breen and Harte, 1993; Yu et al., 1998).

Chromosome translocations characteristically found in human infant leukemia disrupt *MLL* (11q23), generating chimeric proteins between the MLL N terminus and multiple translocation partners that vary substantially (Ayton and Cleary, 2001; Domer et al., 1993; Downing and Look, 1996; Gu et al., 1992; Thirman et al., 1993; Tkachuk et al., 1992). Mice carrying engineered *Mll* translocations develop leukemia (Corral et al., 1996; Forster et al., 2003). Gene expression profiles of infant leukemias bearing *MLL* translocations identified a characteristic gene expression profile that distinguishes this poor prognosis leukemia from other leukemias (Armstrong et al., 2002; Yeoh et al., 2002). Among the upregulated genes were some recognized targets of *MLL* including select *HOX* genes. Deregulated expression of *HOX* genes typifies certain malignancies (Buske and Humphries, 2000; Cillo et al., 2001; Dash and Gilliland, 2001).

Recently, we and others demonstrated that MLL is normally processed at two cleavage sites CS1 (D/GADD) and CS2 (D/GVDD) and that mutation of both sites abolishes the proteolysis (Hsieh et al., 2003; Yokoyama et al., 2002) (Figure 1B). The sequence of the cleavage site is highly conserved in MLL homologs from flies to mammals. MLL cleavage generates N-terminal p320 (N320) and C-terminal p180 (C180) fragments, which heterodimerize to form a stable complex that localizes to a subnuclear compartment. The FYRN domain of N320 directly interacts with the FYRC and SET domains of C180. This dynamic postcleavage association confers stability to N320 and correct nuclear sublocalization of the MLL complex for proper target gene expression (Hsieh et al., 2003).

Site-specific proteolysis is essential in many important biological pathways including the sequential activation of blood coagulation factors (Furie and Furie, 1992), cholesterol-gauged liberation of SREBP from the ER (Brown et al., 2000), ligand-activated cleavage and subsequent release of the intracellular domain of Notch (Brown et al., 2000), maturation of the hedgehog-signaling molecule (Ye and Fortini, 2000), separation of HCF-1 for proper cell cycle regulation (Wilson et al., 1995), and activation of caspases and their subsequent cleavage of death substrates during apoptosis (Thornberry and Lazebnik, 1998). Identification and characterization of the responsible proteases has not only proven critical to understanding such biologic processes but also for developing targeted therapeutics for diseases involving specific pathways. To better understand the regulation of MLL cleavage and hence its effects on target gene expression, we purified the protease responsible for MLL cleavage.

## Results

### Site Specific MLL Cleavage Substrate

Inhibitors to the known classes of proteases showed no substantial inhibition of MLL processing within whole cells, providing the first suggestion that this cleavage activity would prove unique. To facilitate the characterization of the MLL protease, we generated a tractable cleavage substrate. We found that aa 2400–2900 of MLL containing CS1 (aa 2666) and CS2 (aa 2718) are sufficient to recapitulate endogenous MLL cleavage (Figure 1C). Proteolysis of this p75 MLL test substrate at CS1 or CS2 would generate N-terminal Myc-tagged p42 or p47 fragments and C-terminal Flag-tagged p33 or p28, respectively (Figure 1A). The p47 and p28 fragments were

*Correspondence: stanley_korsmeyer@dfci.harvard.edu

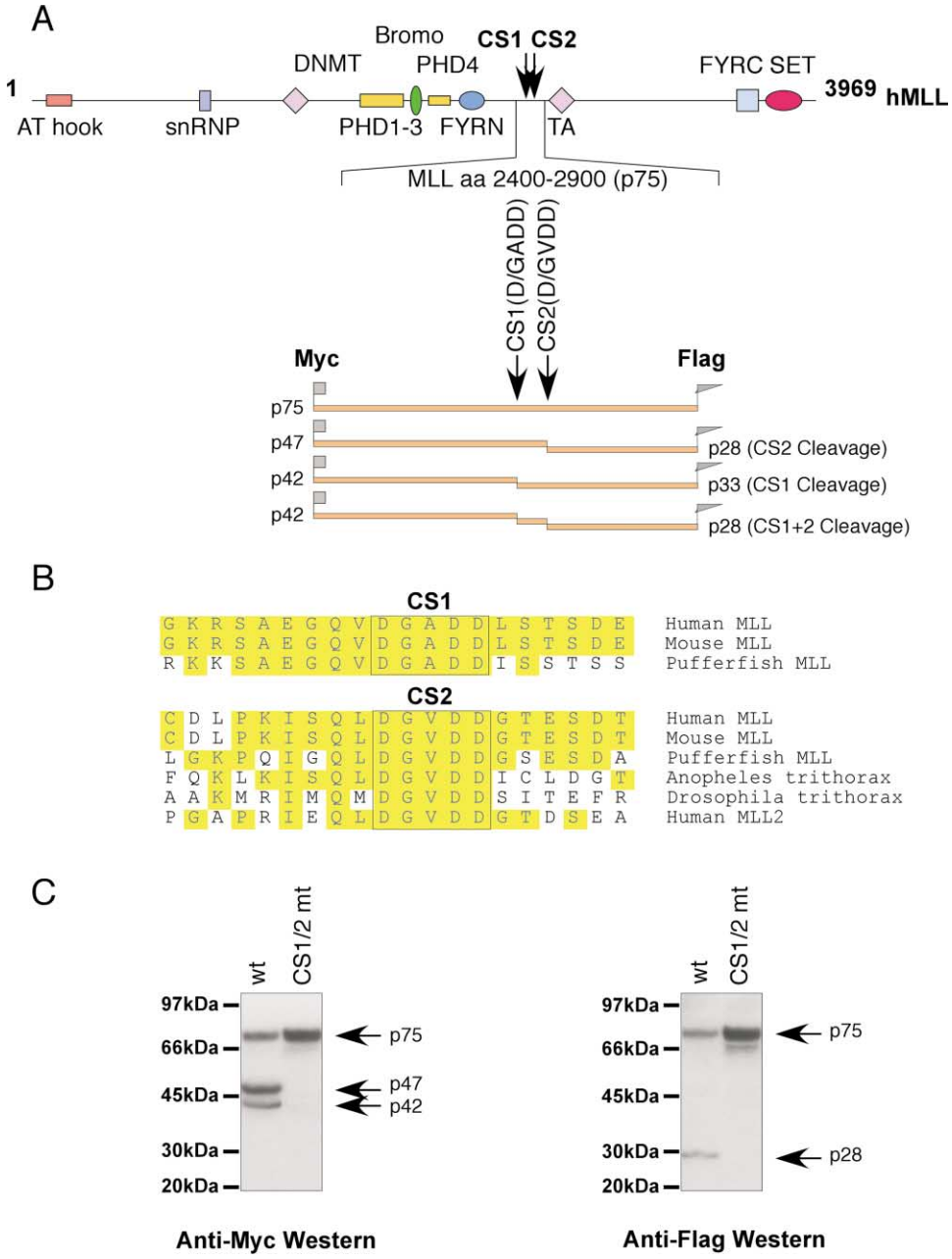


Figure 1. A Cleavage-Site Reporter for MLL Proteolysis

(A) Conserved domain structure of human MLL with cleavage sites (CS1 and CS2). Predicted N-terminal and C-terminal tagged proteolytic products of a 75 kDa wt MLL fragment spanning aa 2400–2900.

(B) Conservation of CS1 (D/GADD) and CS2 (D/GVDD) among MLL family members.

(C) Doubly tagged p75 MLL reporter fragments were transfected into 293T cells and proteolysis by endogenous protease assessed by anti-Myc (left image) or anti-Flag (right image) Western blots. The p47 and p28 cleaved products of wt p75 MLL were most abundant, indicating a preference for CS2 over CS1. The mutant p75 MLL (CS1/2 mt) bearing CS1 (DG2666AA) and CS2 (DGV2718AAA) sequences was not cleaved.

most abundant, indicating processing at CS2 is more efficient than at CS1. This is consistent with our prior observations of the proteolysis of full-length MLL protein (Hsieh et al., 2003). The transfected p75 MLL substrate with mutant CS1/CS2 sites (CS1/2 mt) was not cleaved indicating the specificity of this substrate (Figure 1C). Subcellular fractions derived from 293T cells (a human embryonic kidney cell line) were incubated with in vitro transcribed/translated $^{35}$S-methionine-labeled p75 MLL substrates. The light membrane (LM) fraction displayed the most enzymatic activity (Figure 2A, left image) and did not cleave the mutant fragment (CS1/2 mt) (Figure 2A, right image). Inhibitors of multiple classes of proteases including serine proteases, cysteine proteases,

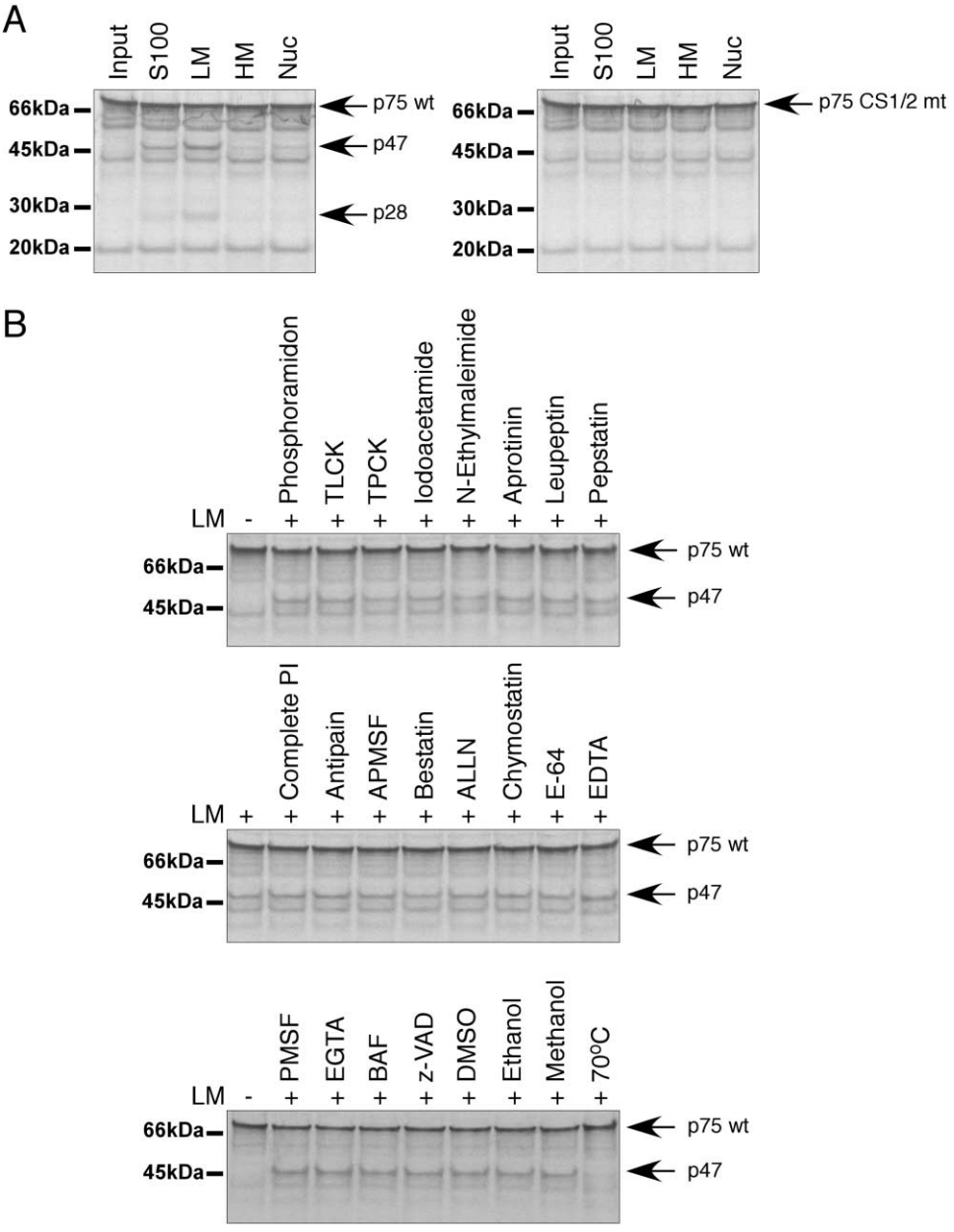


Figure 2. In Vitro Reconstitution of MLL Cleavage Using Subcellular Fractions

(A) The light membrane (LM) subcellular fraction of 293T cells contains enzymatic activity that cleaves the wt but not mutant (CS1/2 mt) p75 MLL reporter, judged by the appearance of p47 and p28 fragments. Lower level activity also exists in the S100 cytosolic fraction.

(B) None of the tested protease inhibitors affected the MLL cleavage activity in the LM fraction. $^{35}$S-methionine-labeled in vitro transcribed/translated p75 MLL substrates were incubated with indicated subcellular fractions and the proteolysis of the reporter fragment analyzed by SDS-PAGE followed by autoradiography.

metalloproteases, acid proteases, and the 26S proteasome, were examined but again showed no substantial inhibition of MLL proteolysis in this fraction enriched for endoplasmic reticulum (Figure 2B). Only heat incubation at 70°C for 30 min inactivated the proteolytic activity in light membranes.

### Purification

The LM fraction possessing the protease activity was subjected to serial column chromatography and the activity followed by an in vitro cleavage assay. Seven chromatographic steps achieved an approximately 200,000-fold enrichment of the proteolytic activity (Figure 3A). The ultimate mono S fractions displaying the highest enzymatic activity were subjected to SDS-PAGE followed by a silver stain (Figure 3B). Polypeptide bands whose presence best correlated with the proteolytic activity were digested with trypsin and subjected to liquid chromatography and tandem mass spectrometry (LC-MS/MS) sequence analysis. Two peptide sequences

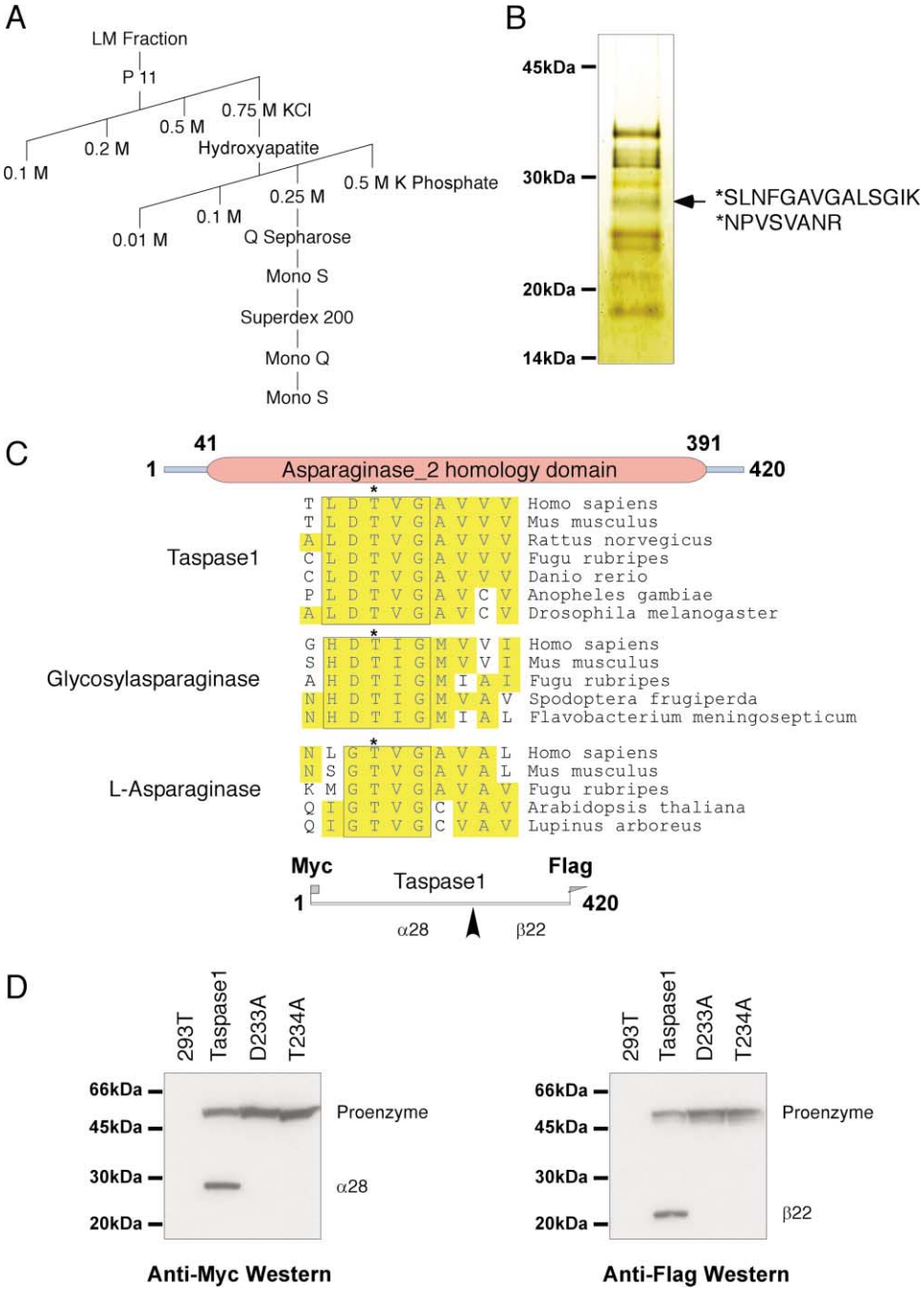


Figure 3. Purification and Cloning of the MLL Cleaving Protease, Taspase1

(A) Schematic of the seven chromatographic steps utilized to enrich the MLL cleaving protease. The enzymatic activity was followed by the in vitro cleavage assay (Figure 2).

(B) Silver-stained SDS-PAGE of the mono S fraction with maximum enzymatic activity. The asterisks denote two peptide sequences obtained from a 28 kDa band whose abundance correlated with enzyme activity (arrow).

(C) Alignment of active sites among Asparaginase_2 family proteins recognizes three distinct conserved motifs that distinguish the Taspase1, glycosylasparaginase and L-asparaginase subfamilies. Conserved active site threonine nucleophiles indicated by asterisks.

(D) Taspase1 was doubly tagged and transiently expressed in 293T cells. Taspase1 is processed from a 50 kDa precursor to an N-terminal 28 kDa α and a C-terminal 22 kDa β subunit in 293T cells (lane 2, both images). Substitution of either aspartate 233 (D233A) or threonine 234 (T234A) with alanine abolished the processing of the Taspase1 proenzyme (lanes 3 and 4, both images).

corresponding to aa 124–137 and aa 138–145 of an uncharacterized open reading frame (orf) present on chromosome 20 (c20orf13) were identified from a gel slice that migrated at ~28 kDa (Figure 3B). The orf predicts a 420 aa protein possessing an Asparaginase_2 homology domain (pfam 01112, interpro 000246) from aa 41–391

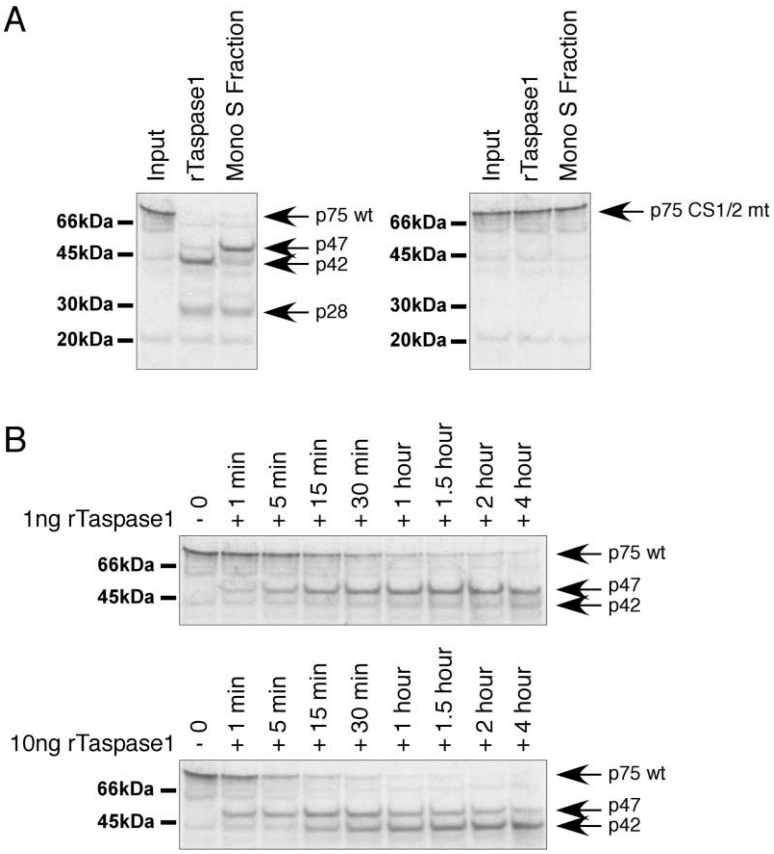


Figure 4. Cleavage of MLL Substrate by Recombinant Taspase1

(A) Purified recombinant Taspase1 (rTaspase1) at 100 ng processes $^{35}$S-methionine-labeled p75 wt MLL substrate (left image) but not p75 CS1/2 mt substrate (right image). Taspase1 processed p75 more completely than the final mono S fraction utilized in Figure 3B (lane 2 and 3, left image).

(B) Taspase1 sequential cleavage of CS1 and CS2. $^{35}$S-methionine labeled p75 wt MLL substrate was incubated with 1 ng or 10 ng rTaspase1 at a final volume of 25 $\mu$l and samples were analyzed over time by SDS-PAGE and autoradiography for cleavage. The more conserved CS2 was cleaved (p47 fragment) before CS1 (p47 fragment) at 1 ng rTaspase1 (upper image). 10-fold more rTaspase1 (10 ng) increased CS1 processing (lower image).

(Figure 3C). Typical proteins that contain this signature motif include L-asparaginase and glycosylasparaginase (Figure 3C). L-asparaginase catalyzes an amide bond hydrolysis to convert L-asparagine to L-aspartate. Glycosylasparaginase is an amidohydrolase, which catalyzes the N-acetylglucosamine-asparagine bond that links oligosaccharides to asparagine. However, no endopeptidase activity had been demonstrated to date among the Asparaginase_2 family enzymes. The characteristics of this MLL cleaving protease (c20orf13) as subsequently detailed here prompt its designation as Taspase1 (*T*hreonine *asp*ar*tase*1). Sequence alignment searches of the public databases identified highly conserved Taspase1 orthologs in fly, mosquito, pufferfish, zebrafish, rat, mouse, and human (Figure 3C, Supplemental Figure S1 available at http://www.cell.com/cgi/content/full/115/3/293/DC1), but not in the nematode, *C. elegans*. The conserved LDTVG motif that surrounds a putative threonine active site is distinct from L-asparaginase and glycosylasparaginase (Figure 3C) suggesting it may have a unique specificity.

### Recombinant Taspase1 Activity

To assess the potential activity of this candidate protease, we expressed and purified recombinant N-terminal His-tagged Taspase1 from *E. coli*. Recombinant Taspase1 (rTaspase1) cleaved the wt p75 MLL but not the CS1/2 mt substrate (Figure 4A). rTaspase1 more efficiently processed p75 MLL to completion based on the predominance of the p42 fragment when compared to the activity in the original mono S fraction (Figure 4A). The p42 fragment results from cleavage at the CS1 (D/GADD) site, which is less conserved than CS2 (D/GVDD) and is also less efficiently processed in vivo (Hsieh et al., 2003). We next examined the differential sensitivity of CS1 and CS2 sites for cleavage by rTaspase1. The p47 fragment appears first, while higher enzyme concentration or more time is needed for the appearance of the p42 fragment indicating a preference of rTaspase1 for CS2 over CS1 (Figure 4B). Thus, the cloned Taspase1's proteolytic characteristics match the pattern of endogenous MLL proteolysis.

### The First *T*hreonine *Asp*ar*tase*, Taspase1

Purification of recombinant N-terminal His-tagged Taspase1 yielded an expected 50 kDa product, a His-tagged 28 kDa $\alpha$ subunit, and a copurified 22 kDa polypeptide (Figure 5A). This 22 kDa polypeptide was subjected to N-terminal Edman degradation analysis, which identified threonine 234 of Taspase1 as the N-terminal amino acid of the apparent 22 kDa $\beta$ subunit (Figure 5A). This represents proteolysis between aspartate 233 and threonine 234 of the 50 kDa putative proenzyme. Similarly when an N-terminal and C-terminal epitope tagged human Taspase1 cDNA was expressed in the human 293 T cell line, the 50 kDa product also appeared to be processed to a 28 kDa N-terminal $\alpha$ subunit and a 22 kDa C-terminal $\beta$ subunit (Figure 3D). This suggested that Taspase1 may be intramolecularly proteolyzed and processed subunits reassembled through a noncovalent association. Amino acid substitution of either aspartate 233 (D233A) or threonine 234 (T234A) to alanine each

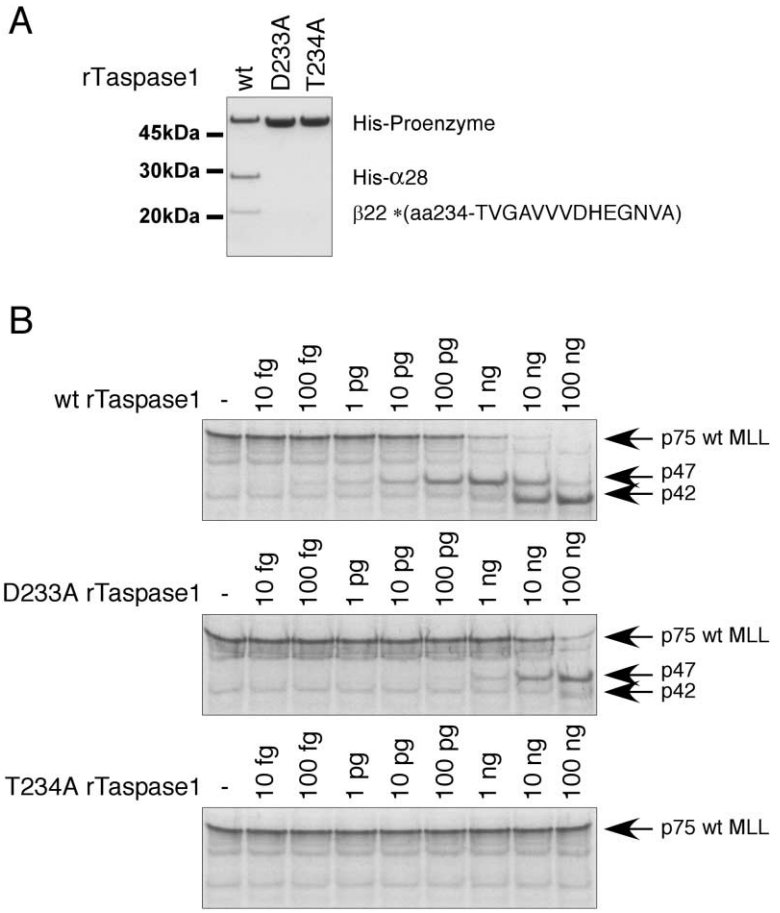


Figure 5. Taspase1 Proenzyme Undergoes Intramolecular Processing to an Active 28 kDa α/22 kDa β Heterodimer

(A) Cobalt column purification of bacterial-expressed, N-terminal His-tagged wt and mutant Taspase1. A 22 kDa fragment copurified with the His-tagged 28 kDa α subunit of Taspase1 (lane 1). Edman degradation of the 22 kDa fragment revealed a 14 amino acid sequence of Taspase1 beginning at threonine 234 (asterisk), which proved to be the 22 kDa β subunit that heterodimerizes with His-α28. Anti-His tag Ab recognized both the 50 kDa His-proenzyme and the processed His-α28 but not the processed β22 of rTaspase1 (data not shown). Alanine substitutions of aspartate 233 (D233A) or threonine 234 (T234A) abolish the intramolecular proteolysis of Taspase1 (lanes 2 and 3).

(B) Purified rTaspase1 in a dose-response assay cleaves $^{35}$S-methionine-labeled p75 wt MLL substrate. The D233A mutant rTaspase1 (middle image) demonstrated an ~1,000-fold decrease in MLL proteolysis. The T234A mutant rTaspase1 demonstrated no enzymatic activity consistent with T234 being the active site nucleophile (lower image).

abolished the intramolecular processing of Taspase1 expressed in *E. coli* (Figure 5A) or in mammalian cells (Figure 3D). However, the D233A mutant retained some residual enzymatic activity, although it was ~1000-fold less efficient than the wt enzyme (Figure 5B, middle image). Conversely, threonine 234 which became the N terminus of the β subunit is absolutely essential for cleavage activity (Figure 5B, lower image). These enzymatic characteristics are similar to properties shared by L-asparaginase and glycosylasparaginase which also demonstrate autoproteolysis of a proenzyme into an active α/β heterodimeric enzyme in which the N-terminal threonine of the β subunit is the active site nucleophile for catalysis (Guan et al., 1996; Liu et al., 1998; Tikkanen et al., 1996; Xu et al., 1999). Thus, to our knowledge, this MLL cleaving protease is the only known endopeptidase within the Asparaginase_2 family with the characteristic of being a *t*hreonine *asp*ar*tase*, Taspase1.

### Proteolysis of MLL In Vivo Requires Taspase1

We next asked whether Taspase1 was required to cleave MLL within mammalian cells. As a model system to test specificity, we coexpressed the p75 MLL substrate reporter together with Taspase1 in 293T cells. Taspase1 resulted in cleavage of wt p75 MLL but not the p75 CS1/2 mt reporter (Figure 6A). Wt Taspase1, but not the T234A mutant Taspase1, enhanced the processing of p75 MLL to the final p42 product (Figures 1A and 6A). Only wt Taspase1, but not the T234A mutant of the nucleophile site demonstrated intramolecular processing into α/β fragments (Figures 3D and 6A). To assess the role of endogenous Taspase1, we designed duplex RNAi against Taspase1, which knocked down the expression of endogenous Taspase1 by ~80% in HeLa cells (Figures 6B and 6C). This reduction of endogenous Taspase1 severely impaired the cleavage of transfected p75 wt MLL reporter as judged by the marked decrease in the processed p47 fragment (11 ± 3% of control RNAi; $n = 3$ experiments; $p < 0.005$) and the disappearance of the p42 fragment (Figure 6B). Taspase1 RNAi knockdown also resulted in a ~50% decrease in the endogenous, processed C180 MLL fragment as well as an increased abundance of full-length p500 MLL (4.3-fold increase over control RNAi; $n = 3$ experiments; $p < 0.005$) (Figure 6C). In contrast, MLL RNAi resulted in the marked reduction of the C180 MLL fragment, but did not increase p500 MLL (Figure 6C). In total, both the in vitro and in vivo cleavage assays confirm the role of Taspase1 in the proper processing of MLL.

### Taspase1 Is Required for Proper *HOX* Gene Expression

Genetic studies in both mice and flies establish that *Mll* and *trithorax* regulate *Hox* and *homeotic gene* expression, respectively (Mazo et al., 1990; Yu et al., 1995). Mice heterozygous for an *Mll-exon3$^{LacZ}$* disruption dem-

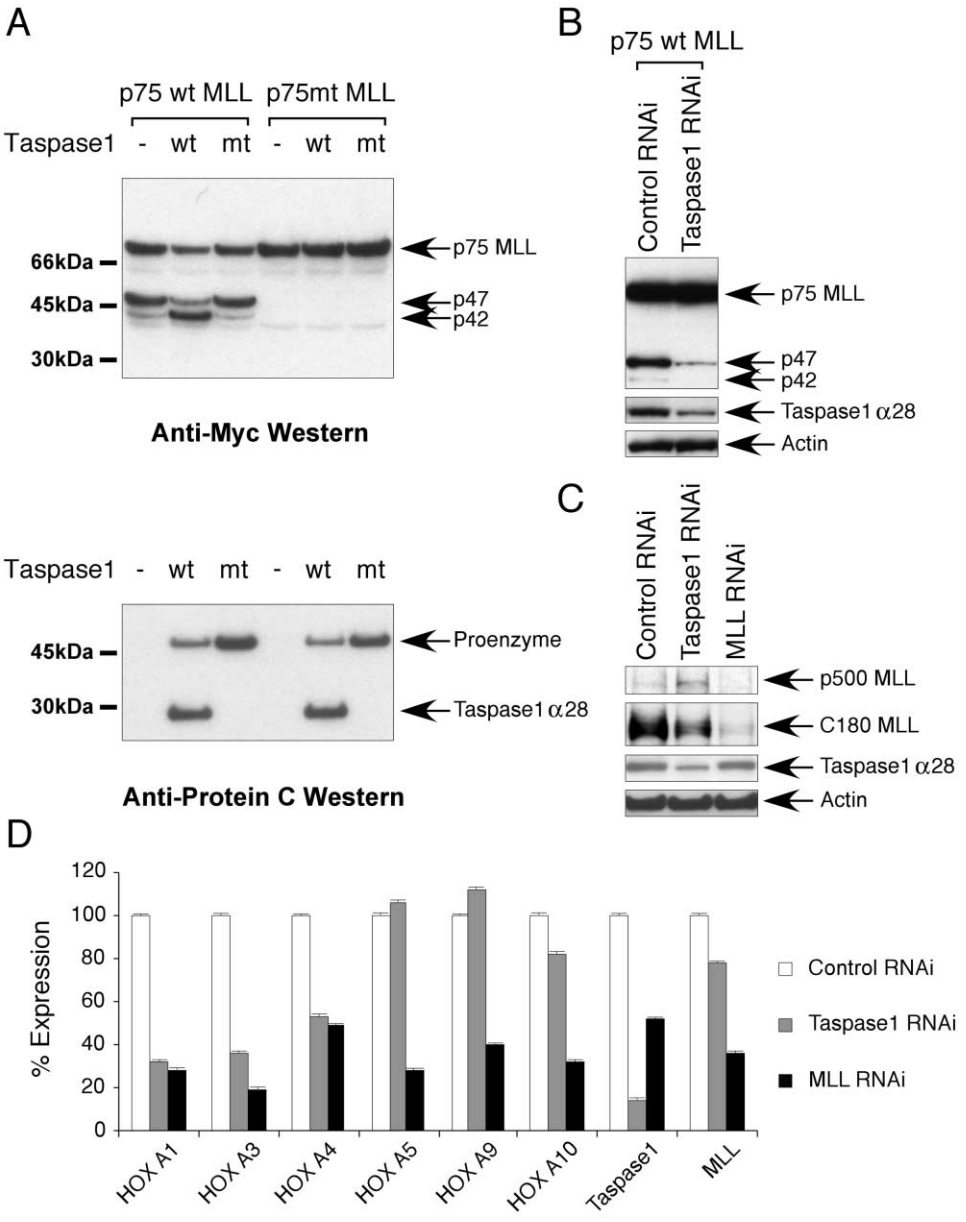


Figure 6. Taspase1 Is Required for Endogenous MLL Cleavage and Proper *HOX* Gene Regulation

(A) Constructs expressing wt or mt p75 MLL substrate (as in Figure 1C) were cotransfected with wt or mt (T234A) Taspase1 bearing an N-terminal protein-C tag into 293T cells. The N-terminal cleavage products of p75 MLL were detected by an anti-Myc Western blot. The cleavage of wt p75 MLL substrate by endogenous protease is shown in lane 1, upper image. Transfection of wt Taspase1 enhanced complete processing of p75 MLL processing, judged by the decrease of both p75 and p47, and the increase of the p42 fragment (lanes 2, upper image). No enhancement was observed with transfected T234A mutant (lane 3, upper image). The intramolecular proteolysis of expressed wt but not the T234A mutant Taspase1 is demonstrated by antiprotein C Western blot (lower image).

(B) RNAi-mediated knockdown of endogenous Taspase1 impairs the processing of p75 wt MLL. HeLa cells were treated with indicated duplex RNAi for 48 hr before introduction of the p75 wt MLL expression construct and analysis of cleavage 24 hr later. Taspase1 knockdown leads to a marked decrease in the p47 and the disappearance of the p42 fragment (lane 2).

(C) RNAi-mediated knockdown of endogenous Taspase1 results in increased abundance of unprocessed p500 MLL and diminished processed C180 MLL (lane 2). The indicated duplex RNAi was introduced into HeLa cells and the expression of Taspase1 α28, p500 MLL, C180 MLL, and actin were detected by the respective Abs. MLL RNAi markedly reduced the expression C180 MLL (lane 3). Control duplex RNAi had no effect (lane 1).

(D) Altered *HOX A* gene expression in Taspase1 knockdown cells. Quantitative RT-PCR analysis of indicated gene expression of RNAi-treated HeLa cells is shown as a bar graph. The expression of individual genes is normalized to GAPDH levels for individual samples. The % expression of an indicated gene is calculated in comparison to a control duplex RNAi set at 100%. Values shown are mean ± 1 SD of four independent experiments.

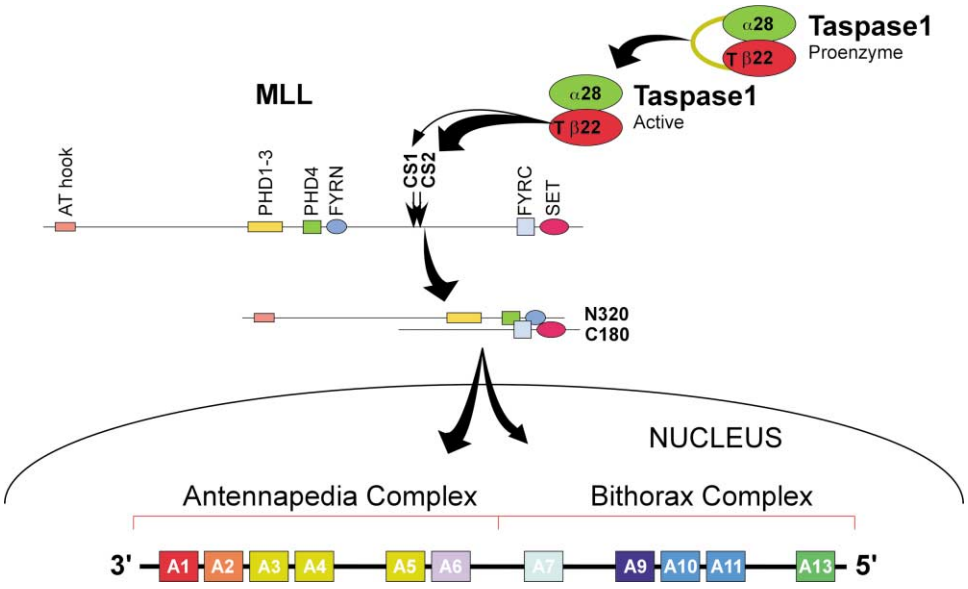


Figure 7. A Schematic Model Depicts the Intramolecular Proteolysis of Taspase1 Followed by MLL Processing Required for Proper *HOX* Gene Expression

onstrate haploinsufficiency with bidirectional homeotic transformations and shifted anterior boundaries of several *Hox* genes (Yu et al., 1995). *Mll*$^{-/-}$ deficient embryos and mouse embryonic fibroblasts (MEFs) demonstrate *Mll* is required for the maintenance of selected *Hox* gene expression (Hanson et al., 1999; Yu et al., 1995, 1998). As a first assessment of whether reduced Taspase1 activity would alter gene expression, we examined the gene expression profile of HeLa cells treated with the Taspase1 RNAi versus a control RNAi (Figure 6D). Initial analysis of Affymetrix (HG-U133A) oligonucleotide array-based RNA profiles indicated diminished expression of selected *HOX* genes (data not shown). Consequently, we used a quantitative RT-PCR approach to determine the relative expression of genes across the *HOX A* cluster. Of note, the knockdown of *Taspase1* diminished the expression of the 3′ located and "earlier" expressed genes in the *HOX A* cluster (*A1, A3*, and *A4*), but not those genes located more 5′ and expressed "later" during embryonic development (*A5*, *A9*, and *A10*) (Figure 6D). This selected attenuation contrasts with the global decrease in expression of most *HOX A* genes (*A1* to *A10*) in cells with *MLL* knocked down (Figure 6D). These data support the importance of Taspase1 in the correct expression of the early *HOX A* gene cluster (equivalent to the *ANT-C* cluster of *Drosophila*) (Figure 7).

## Discussion

We utilized highly conserved cleavage sequences within MLL to purify and clone a novel protease responsible for the routine proteolysis of MLL. Taspase1 (*t*hreonine *asp*artase 1) is an endopeptidase within a family of enzymes possessing an Asparaginase_2 homology domain. Other members present in both prokaryotes and eukaryotes include the amidohydrolases, L-asparaginase in intermediary amino acid metabolism and glycosylasparaginase. Glycosylasparaginase participates in the ordered degradation of N-linked glycoproteins by cleaving Asn-GlcNAc linkages that join oligosaccharides to proteins. Taspase1-mediated cleavage of MLL follows distinct aspartate residues suggesting Taspase1 evolved from hydrolyzing asparagine and glycosylasparagine to recognize a conserved peptide motif with an aspartate at the P1 position. Taspase1 also belongs to a larger superfamily of diverse enzymes that are Ntn-hydrolases (*N-t*erminal *n*ucleophiles) (Brannigan et al., 1995). All members of this family are activated autocatalytically, displaying an N terminally located catalytic nucleophile. The nascent single polypeptide is cleaved into nonidentical α and β subunits, which assemble into a catalytically active αββα heterodimeric structure (Oinonen and Rouvinen, 2000). Glycosylasparaginase is a well-characterized prototype that self-catalyses a peptide bond rearrangement. Of note, the threonine that provides the nucleophile for both autoproteolysis and hydrolase activity as well as the immediately upstream aspartate residue are conserved in Taspase1. The threonine nucleophile of glycosylasparaginase appears to attack this immediately upstream amide bond through an N->O acyl shift to catalyze this peptide bond cleavage (Xu et al., 1999). In the Ntn-hydrolase superfamily the nucleophile always originates from the N terminus of the β subunit. The complete dependence of Taspase1 upon thr234 for both its intramolecular proteolysis to generate an α/β heterodimer as well as its proteolytic activity for MLL substrates supports its role as the active nucleophile and the inclusion of Taspase1 in this family.

Caspases and Granzyme B are well-characterized proteases that also cleave peptide bonds after a P1 aspartate (Odake et al., 1991; Thornberry et al., 1992). However, they belong to structurally distinct, large families of proteases that use cysteine or serine as the active nucleophile. Among the known proteases, the β subunit of the 20S proteasome, the β subunit of the archea proteasome, and the catalytic subunit of *E. coli* HslV are

known threonine endopeptidases (Bochtler et al., 1999; Seemuller et al., 1995). While the 20S proteasome may be responsible for site-specific cleavage of precursor proteins such as p105 NFkB (Palombella et al., 1994), it predominantly participates in protein degradation (Hochstrasser, 1996).

MLL family proteins have architecturally positioned domains with distinct biochemical properties (Figure 1A and Supplemental Figure S2 available at *Cell* website). Vertebrate MLL has two conserved cleavage sites, CS1 and CS2 (Hsieh et al., 2003). Of note, other homologs including invertebrate *trithorax* and vertebrate *MLL2* only possess the more conserved CS2 site. These cleavage sites are strategically positioned such that proteolyzed MLL fragments form a heterodimeric complex utilizing the FYRN domain of N320 and the FYRC plus SET domains of the C180 fragment (Hsieh et al., 2003). Consequently, the postcleavage reassembly of MLL fragments would be predicted to bring four domains (i.e., PHD4, FYRN, FYRC, and SET) in close proximity. This postcleavage juxtaposition is reminiscent of the contiguous four-domain cluster, PHD4-FYRN-FYRC-SET, found at the C terminus of more evolutionarily distant MLL homologs (Supplemental Figure S2 available at *Cell* website). This latter set of proteins is present in all *Metazoa* including T12D8.1 of *C. elegans* and *trr* (*trithorax-related*) of *Drosophila, ALR* (*ALL1 Related*) and *MLL3* of *Homo sapiens* (Supplemental Figure S2 available at *Cell* website). These proteins with clustered domains do not display the cleavage sequence and recombinant MLL3 was not proteolyzed (data not shown). The proximity of these four domains may be essential for a common biochemical property of this protein family. Higher organisms with complex body plans require additional levels of regulation for specification programs including those mediated by *Hox* genes. *Chordata* and *Arthropoda* possess MLL homologs with motifs proteolyzed by the coevolved Taspase1, which distinguishes them from the more ancient four-domain cluster.

Deregulated expression of selected *HOX* genes accompanies some human malignancies including leukemias bearing chromosomal translocations that result in MLL-fusion proteins (Armstrong et al., 2002; Yeoh et al., 2002). Recently, the transformation of myeloid progenitors mediated by an MLL-ENL fusion protein was shown to depend on selected downstream *Hox* genes (Ayton and Cleary, 2003). The discovery of Taspase1 provides a unique opportunity to probe the role of MLL cleavage in both physiologic and pathologic settings. Finally, our data warrant development of a Taspase1 inhibitor that would be predicted to impair expression of selected *HOX* genes and might prove a rational approach to cancer therapy.

*Homeotic* genes of the invertebrates and *Hox* genes of the vertebrates encode a family of related transcription-regulatory proteins that play essential roles in implementing developmental patterns and conferring critical positional information (Krumlauf, 1994). The initiation of *Hox* gene expression is controlled by transiently expressed segmentation gene products. However, the maintenance of appropriate *Hox* expression patterns is conferred by the positive regulatory *trithorax* group (*trx-G*) and the negative regulatory *polycomb* group (*Pc-G*), which heavily utilize epigenetic mechanisms (Pirrotta, 1998). *Mll* loss-of-function studies indicate it is required for maintaining *Hox* expression during early embryogenesis (Yu et al., 1995). Moreover, the axial-skeletal transformations and altered *Hox* patterns of *Mll*-deficient mice were normalized when *PcG* member *Bmi*-1 was deleted (Hanson et al., 1999). The CS2 (D/GVDD) cleavage site in *Drosophila* TRX resides within a 281aa region deleted in a mutant *trx* allele, $trx^{E3}$ (Mazo et al., 1990). Null mutations of *trx* fail to maintain the expression of *homeotic* genes of both the *bithorax* (*BX-C*) and *antennapedia* (*ANT-C*) complexes (Breen, 1999). Whereas, $trx^{E3}$ displays selectively reduced expression of *ANT-C* but not *BX-C* genes (Sedkov et al., 1994). Subsequent studies using antibodies recognizing different regions of TRX support the CS2 sequences serving as a proteolytic site (Kuzin et al., 1994). Similarly, our RNAi knockdown of Taspase1 in mammalian cells suggests that cleavage of MLL is important in maintaining expression of the "early" 3′ located *HOX A* genes (*HOX A1, A3, A4* analogous to *ANT-C)* but not the "late" more 5′ located *HOX A* genes (*HOX A9, A10* analogous to *BX-C*) (Figure 7). In contrast, RNAi knockdown of *MLL* and our *Mll*-null cells display more global defects in *Hox* gene expression similar to *trx*-null alleles (data not shown). Perhaps the earliest expressed 3′ *HOX* genes are most dependent upon processed MLL, raising the possibility that the precise multiprotein complexes assembled by MLL might vary along the *HOX* gene cluster. A role for Taspase1-mediated cleavage of MLL in proper *HOX* gene expression suggests *Taspase1* coevolved with *MLL/trithorax* at the time when *Arthropoda* and *Chordata* emerged from *Metazoa*, enabling the more complex axial skeletal and developmental plans of higher organisms.

## Experimental Procedures

### Plasmid Construction and Antibody Production

PCR fragments consisting of MLL aa 2400–2900 derived from either wild-type or noncleavable MLL mt (CS1/CS2 mt) were inserted into a Myc/Flag doubly tagged eukaryotic expression vector for transient transfection assays. These constructs also contain a 5′ T7 promoter for generating in vitro transcription/translation product of $^{35}$S-methionine-labeled p75 MLL substrates. Full-length Taspase1 was cloned from 293T cell cDNA and inserted into the Myc/Flag doubly tagged expression vector, a protein C-tagged vector, and a His-tagged bacteria expression vector, ET15b (Novagen). Taspase1 mutants were generated using QuikChange site-directed mutagenesis kit (Stratagene). Rabbit anti-Taspase1 polyclonal antibody was raised against aa 7–212 of purified recombinant human Taspase1. Transient transfection, in vitro transcription/translation, $^{35}$S-methionine labeling, and immunoblot assays were performed as previously described (Hsieh et al., 2003).

### In Vitro Cleavage Assays

$^{35}$S-methionine-labeled MLL substrate was incubated with 2 μl of indicated subcellular fractions or specified amounts of rTaspase1 in cleavage buffer (100 mM HEPES [pH 7.9], 5 mM $MgCl_2$, 20 mM KCl, 5 mM DTT, and 10% sucrose) at a final volume of 25 μl for 1 hr at 37°C or indicated periods of time. Protease inhibitors utilized include 8.5 μM phosphoramidon, 100 μM TLCK, 100 μM TPCK, 5 mM iodoacetamide, 5 mM N-ethylmaleimide, 0.3 μM aprotinin, 100 μM leupeptin, 1 μM pepstatin, 1× complete protease inhibitor cocktail (Roche), 100 μM anti-pain, 100 μM APMSF, 10 μM bestatin, 25 μM ALLN, 100 μM chymostatin, 10 μM E-64, 5 mM EDTA, 1 mM PMSF, 1 mM EGTA, 50 μM BAF (Boc-Aspartyl-FMK), and 50 μM z-VAD (z-VAD-FMK).

### Purification and LC-MS/MS

Human 293T cells from one hundred 15 cm dishes were collected and incubated in hypotonic buffer (10 mM HEPES [pH 7.9], 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM PMSF, 1 mM EDTA, 1 mM EDTA, 1 mM EGTA, and protease inhibitor cocktail [Roche]) for 15 min on ice. Supplemented protease inhibitors were omitted during the initial characterization of MLL cleaving protease. Cells were homogenized using a glass dounce and the homogenized cellular extract was subjected to centrifugal fractionation. Nuclei (Nuc) and unbroken cells were twice separated at 700 g for 10 min. The heavy membrane (HM) fraction pellet was collected after two centrifugations of the supernatant at 7000 g for 10 min. The resulting supernatant was centrifuged at 100,000 g for 30 min to yield the light membrane pellet (LM) and final soluble fraction (S100). Proteins were solubilized in buffer A (20 mM HEPES [pH 7.9], 100 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM PMSF, 1 mM DTT, 1 mM EDTA, 1 mM EGTA, 0.1% Tween 20, and 10% glycerol) with additional 0.5% of Tween 20. Solubilized LM fraction was applied to a P11 column and the bound protease was eluted with gradients of KCl. Positive fractions were collected and dialyzed against buffer B (10 mM HEPES [pH 7.9], 100 mM KCl, 1 mM $MgCl_2$, 10 $\mu$M $CaCl_2$, 0.2 mM PMSF, 1 mM DTT, 1 mM EDTA, 1 mM EGTA, 10% glycerol, 0.1% Tween 20, and 10 mM potassium phosphate [pH 7.9]) and applied to a hydroxyapatite column. Elution was performed with a phosphate gradient and the protease positive fractions were dialyzed against buffer A before loading onto indicated chromatographic columns. LC-MS/MS was performed by the Taplin Biological Mass Spectrometry Facility at the Harvard Medical School.

### Recombinant Enzyme and Edman Degradation Analysis

His-tagged Taspase1 was expressed in BL21(DE3) cells and purified with a TALON column (Clontech). N-terminal protein sequencing was performed by Molecular Biology Core Facility at the Dana-Farber Cancer Institute.

### RNAi, Reverse Transcription, and Quantitative PCR

HeLa cells were transfected with indicated duplex RNAi (Dharmacon) using oligofectamine (Invitrogen). Double-stranded ribooligonucleotides with overhanging 3′ deoxy TT were prepared to target mRNAs of either hTaspase1 (GACUCACAUUUCAAGACUU) or hMLL (GAAGUCAGAGUGCGAAGUC). Cells harvested 72 hr after transfection were either lysed in RIPA buffer for immunoblots or with Trizol (Invitrogen) for RNA purification using RNeasy (Qiagen). Transfection of p75 wt MLL reporter substrate was performed 48 hr following individual RNAi knockdown in HeLa cells and the cellular extracts were harvested 24 hr later for immunoblot analysis. Reverse transcription were performed with oligo-dT primers using Superscript II (Invitrogen). Quantitative PCR was performed in triplicate using indicated gene specific primers (Supplemental Data available at *Cell* website) with SYBR green (PE biosystems) on the ABI Prism 7700 sequence detection system.

### Acknowledgments

We thank C. Walsh for critical advice; A. Lane and T. Ley for helpful discussions; and E. Smith for preparation of the manuscript. This work is supported in part by NIH grants CA102594 and CA68484.

Received: July 28, 2003
Revised: September 25, 2003
Accepted: October 3, 2003
Published: October 30, 2003

### References

Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R., and Korsmeyer, S.J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. *30*, 41–47.

Ayton, P.M., and Cleary, M.L. (2001). Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. Oncogene *20*, 5695–5707.

Ayton, P.M., and Cleary, M.L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes Dev. *17*, 2298–2307.

Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., and Huber, R. (1999). The proteasome. Annu. Rev. Biophys. Biomol. Struct. *28*, 295–317.

Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R., and Murzin, A.G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. Nature *378*, 416–419.

Breen, T.R. (1999). Mutant alleles of the Drosophila trithorax gene produce common and unusual homeotic and other developmental phenotypes. Genetics *152*, 319–344.

Breen, T.R., and Harte, P.J. (1993). Trithorax regulates multiple homeotic genes in the bithorax and antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. Development *117*, 119–134.

Brown, M.S., Ye, J., Rawson, R.B., and Goldstein, J.L. (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell *100*, 391–398.

Buske, C., and Humphries, R.K. (2000). Homeobox genes in leukemogenesis. Int. J. Hematol. *71*, 301–308.

Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. (2001). Homeobox genes in normal and malignant cells. J. Cell. Physiol. *188*, 161–169.

Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G., and Rabbitts, T.H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. Cell *85*, 853–861.

Dash, A., and Gilliland, D.G. (2001). Molecular genetics of acute myeloid leukaemia. Best Pract. Res. Clin. Haematol. *14*, 49–64.

Domer, P.H., Fakharzadeh, S.S., Chen, C.S., Jockel, J., Johansen, L., Silverman, G.A., Kersey, J.H., and Korsmeyer, S.J. (1993). Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. Proc. Natl. Acad. Sci. USA *90*, 7884–7888.

Downing, J.R., and Look, A.T. (1996). MLL fusion genes in the 11q23 acute leukemias. Cancer Treat. Res. *84*, 73–92.

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol. Cell. Biol. *21*, 2249–2258.

Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M., and Diaz, M.O. (2001). Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. Mol. Cell. Biol. *21*, 3589–3597.

Forster, A., Pannell, R., Drynan, L.F., McCormack, M., Collins, E.C., Daser, A., and Rabbitts, T.H. (2003). Engineering de novo reciprocal chromosomal translocations associated with Mll to replicate primary events of human cancer. Cancer Cell *3*, 449–458.

Furie, B., and Furie, B.C. (1992). Molecular and cellular biology of blood coagulation. N. Engl. J. Med. *326*, 800–806.

Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C.M., and Canaani, E. (1992). The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. Cell *71*, 701–708.

Guan, C., Cui, T., Rao, V., Liao, W., Benner, J., Lin, C.L., and Comb, D. (1996). Activation of glycosylasparaginase. Formation of active N-terminal threonine by intramolecular autoproteolysis. J. Biol. Chem. *271*, 1732–1737.

Hanson, R.D., Hess, J.L., Yu, B.D., Ernst, P., van Lohuizen, M., Berns, A., van der Lugt, N.M., Shashikant, C.S., Ruddle, F.H., Seto, M., and Korsmeyer, S.J. (1999). Mammalian trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. Proc. Natl. Acad. Sci. USA *96*, 14372–14377.

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. Annu. Rev. Genet. *30*, 405–439.

Hsieh, J.J., Ernst, P., Erdjument-Bromage, H., Tempst, P., and Korsmeyer, S.J. (2003). Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol. Cell. Biol. *23*, 186–194.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Kuzin, B., Tillib, S., Sedkov, Y., Mizrokhi, L., and Mazo, A. (1994). The Drosophila trithorax gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. Genes Dev. *8*, 2478–2490.

Liu, Y., Guan, C., and Aronson, N.N., Jr. (1998). Site-directed mutagenesis of essential residues involved in the mechanism of bacterial glycosylasparaginase. J. Biol. Chem. *273*, 9688–9694.

Mazo, A.M., Huang, D.H., Mozer, B.A., and Dawid, I.B. (1990). The trithorax gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains. Proc. Natl. Acad. Sci. USA *87*, 2112–2116.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol. Cell *10*, 1107–1117.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M., and Canaani, E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol. Cell *10*, 1119–1128.

Odake, S., Kam, C.M., Narasimhan, L., Poe, M., Blake, J.T., Krahenbuhl, O., Tschopp, J., and Powers, J.C. (1991). Human and murine cytotoxic T lymphocyte serine proteases: subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. Biochemistry *30*, 2217–2227.

Oinonen, C., and Rouvinen, J. (2000). Structural comparison of Ntn-hydrolases. Protein Sci. *9*, 2329–2337.

Palombella, V.J., Rando, O.J., Goldberg, A.L., and Maniatis, T. (1994). The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B. Cell *78*, 773–785.

Pirrotta, V. (1998). Polycombing the genome: PcG, trxG, and chromatin silencing. Cell *93*, 333–336.

Sedkov, Y., Tillib, S., Mizrokhi, L., and Mazo, A. (1994). The bithorax complex is regulated by trithorax earlier during Drosophila embryogenesis than is the Antennapedia complex, correlating with a bithorax-like expression pattern of distinct early trithorax transcripts. Development *120*, 1907–1917.

Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R., and Baumeister, W. (1995). Proteasome from Thermoplasma acidophilum: a threonine protease. Science *268*, 579–582.

Thirman, M.J., Gill, H.J., Burnett, R.C., Mbangkollo, D., McCabe, N.R., Kobayashi, H., Ziemin-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A.A., et al. (1993). Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. N. Engl. J. Med. *329*, 909–914.

Thornberry, N.A., and Lazebnik, Y. (1998). Caspases: enemies within. Science *281*, 1312–1316.

Thornberry, N.A., Bull, H.G., Calaycay, J.R., Chapman, K.T., Howard, A.D., Kostura, M.J., Miller, D.K., Molineaux, S.M., Weidner, J.R., Aunins, J., et al. (1992). A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. Nature *356*, 768–774.

Tikkanen, R., Riikonen, A., Oinonen, C., Rouvinen, R., and Peltonen, L. (1996). Functional analyses of active site residues of human lysosomal aspartylglucosaminidase: implications for catalytic mechanism and autocatalytic activation. EMBO J. *15*, 2954–2960.

Tkachuk, D.C., Kohler, S., and Cleary, M.L. (1992). Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691–700.

Wilson, A.C., Peterson, M.G., and Herr, W. (1995). The HCF repeat is an unusual proteolytic cleavage signal. Genes Dev. *9*, 2445–2458.

Xia, Z.B., Anderson, M., Diaz, M.O., and Zeleznik-Le, N.J. (2003). MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc. Natl. Acad. Sci. USA *14*, 8342–8347.

Xu, Q., Buckley, D., Guan, C., and Guo, H.C. (1999). Structural insights into the mechanism of intramolecular proteolysis. Cell *98*, 651–661.

Ye, Y., and Fortini, M.E. (2000). Proteolysis and developmental signal transduction. Semin. Cell Dev. Biol. *11*, 211–221.

Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., et al. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell *1*, 133–143.

Yokoyama, A., Kitabayashi, I., Ayton, P.M., Cleary, M.L., and Ohki, M. (2002). Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. Blood *100*, 3710–3718.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505–508.

Yu, B.D., Hanson, R.D., Hess, J.L., Horning, S.E., and Korsmeyer, S.J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc. Natl. Acad. Sci. USA *95*, 10632–10636.

Zeleznik-Le, N.J., Harden, A.M., and Rowley, J.D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc. Natl. Acad. Sci. USA *91*, 10610–10614.